



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 9/26, 15/56, C07K 19/00	A1	(11) International Publication Number: WO 98/16633 (43) International Publication Date: 23 April 1998 (23.04.98)
(21) International Application Number: PCT/DK97/00448 (22) International Filing Date: 13 October 1997 (13.10.97) (30) Priority Data: 1130/96 11 October 1996 (11.10.96) DK (71) Applicant (for all designated States except US): NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsværd (DK). (72) Inventors; and (75) Inventors/Applicants (for US only): BJØRNVAD, Mads [DK/DK]; Novo Nordisk a/s, Novo Allé, DK-2880 Bagsværd (DK). PEDERSEN, Sven [DK/DK]; Novo Nordisk a/s, Novo Allé, DK-2880 Bagsværd (DK). SCHULEIN, Martin [DK/DK]; Novo Nordisk a/s, Novo Allé, DK-2880 Bagsværd (DK). BISGÅRD-FRANTZEN, Henrik [DK/DK]; Novo Nordisk a/s, Novo Allé, DK-2880 Bagsværd (DK). (74) Common Representative: NOVO NORDISK A/S; Novo Allé, DK-2880 Bagsværd (DK).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i>
(54) Title: ALPHA-AMYLASE FUSED TO CELLULOSE BINDING DOMAIN, FOR STARCH DEGRADATION (57) Abstract The invention relates to a starch conversion method wherein the starch substrate is treated in aqueous medium with an CBD/enzyme hybrid. Further, the invention also relates to an isolated DNA sequence encoding a stable CBD/enzyme hybrid, a DNA construct comprising said DNA sequence of the invention, an expression vector comprising the DNA sequence of the invention, and a CBD/enzyme hybrid.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakistan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

ALPHA-AMYLASE FUSED TO CELLULOSE BINDING DOMAIN, FOR STARCH DEGRADATION

FIELD OF THE INVENTION

The present invention relates, *inter alia*, to the use of a hybrid between a carbohydrate-binding domain ("CBD") and an enzyme of a type employed in industrial starch processing [notably starch processing for the production (*vide infra*) of sweeteners, particularly glucose- and/or fructose-containing syrups], especially an amylolytic enzyme, such as an α -amylase employed in a so-called "starch liquefaction" process (*vide infra*) in which starch is degraded (often termed "dextrinized") to smaller oligo- and/or polysaccharide fragments, or a debranching enzyme (such as an isoamylase or a pullulanase) employed to debranch amylopectin-derived starch fragments in connection with the so-called "saccharification" process (*vide infra*) which is normally carried out after the liquefaction stage. The invention also relates to hybrid enzyme consisting of a CBD-linker-enzyme.

BACKGROUND OF THE INVENTION

As indicated above, the present invention is of particular value in the field of starch processing (starch conversion). Conditions for conventional starch conversion processes and for liquefaction and/or saccharification processes are described in, e.g., US 3,912,590 and in EP 0,252,730 and EP 0,063,909.

Production of sweeteners from starch:

A "traditional" process for the production of glucose- and fructose-containing syrups from starch normally consists of three consecutive enzymatic processes, viz. a liquefaction process followed by a saccharification process and (for production of fructose-containing syrups) an isomerization process. During the liquefaction process, starch (initially in the form starch suspension in aqueous medium) is degraded to dextrans (oligo- and polysaccharide fragments of starch) by an α -amylase [EC 3.2.1.1; e.g. TermamylTM (*Bacillus licheniformis* α -amylase), available from Novo Nordisk A/S, Bagsvaerd, Denmark], typically at pH

values between 5.5 and 6.2 and at temperatures of 95-160°C for a period of approximately 2 hours. In order to ensure optimal enzyme stability under these conditions, approximately 1mM of calcium (ca. 40 ppm free calcium ions) is typically added to the
5 starch suspension.

After the liquefaction process the dextrins are converted into dextrose (D-glucose) by addition of a glucoamylase (amyloglucosidase, EC 3.2.1.3; e.g. AMG™, from Novo Nordisk A/S) and, typically, a debranching enzyme, such as an isoamylase (EC
10 3.2.1.68) or a pullulanase (EC 3.2.1.41; e.g. Promozyme™, from Novo Nordisk A/S). Before this step the pH of the medium is normally reduced to a value below 4.5 (e.g pH 4.3), maintaining the high temperature (above 95°C), and the liquefying α -amylase activity is thereby denatured. The temperature is then normally
15 lowered to 60°C, and glucoamylase and debranching enzyme are added. The saccharification process is normally allowed to proceed for 24-72 hours.

After completion of the saccharification stage, the pH of the medium is increased to a value in the range of 6-8, preferably pH
20 7.5, and calcium ions are removed by ion exchange. The resulting syrup (dextrose syrup) may then be converted into high fructose syrup using, e.g., an immobilized "glucose isomerase" (xylose isomerase, EC 5.3.1.5; e.g. Sweetzyme™, from Novo Nordisk A/S).

A number of improvements in the properties of enzymes
25 currently employed in starch conversion processes would be desirable. With respect to starch liquefaction, employing liquefying α -amylases, at least 3 improvements could be envisaged and are outlined below; each of these could be regarded as an individual benefit, although any combination (e.g. 1+2, 1+3, 2+3
30 or 1+2+3) could advantageously be employed:

Improvement 1.

Reduction of the Ca^{2+} dependency of the liquefying α -amylase.

Addition of free calcium (calcium ion) is required to ensure
35 adequately high stability of α -amylases currently employed for starch liquefaction, but the presence of calcium ions in the

medium at the isomerization stage results in strong inhibition of the activity of the glucoseisomerase employed therein. It is therefore necessary either to reduce the calcium ion content of the medium, by means of an expensive unit operation (e.g. ion exchange), to a level below about 3-5 ppm of free calcium, or to minimize the inhibitory effect of calcium in some other manner, e.g. by addition, after the saccharification stage, to the medium of magnesium ions in a amount sufficient to adequately "out-compete" binding of calcium to the glucoseisomerase. Significant savings could be achieved if the liquefaction process could be performed without addition of calcium ions, thereby eliminating the need for subsequent, expensive remedial unit operations to remove calcium or minimize the inhibitory effect thereof.

To achieve this, an α -amylolytic enzyme which is stable and highly active at low concentrations of free calcium (< 40 ppm) is required. Such an enzyme should preferably have a pH optimum at a pH in the range of 4.5-6.5, more preferably in the range of 4.5-5.5.

20 Improvement 2.

Reduction of formation of unwanted Maillard products.

The extent of formation of unwanted Maillard products during the liquefaction process is dependent on the pH. Low pH favours reduced formation of Maillard products. It would thus be desirable to be able to lower the process pH from around pH 6.0 to a value around pH 4.5; unfortunately, all commonly known, thermostable liquefying α -amylases are not very stable at low pH (i.e. pH < 6.0) and their specific activity is generally low.

Achievement of the above-mentioned goal requires the availability of an α -amylolytic enzyme which is stable at a pH in the range of 4.5-5.5, and which preferably maintains a high specific activity.

Improvement 3.

35 Reduced influence of the liquefying α -amylase on the saccharification process.

It has been reported previously (US patent 5,234,823) that

when saccharifying with *A. niger* glucoamylase and *B. acidopullu-*
lyticus pullulanase, the presence of residual α -amylase activity
remaining after the liquefaction process can lead to lower yields
of dextrose if the α -amylase is not inactivated before the
5 saccharification stage. As already mentioned (*vide supra*), this
inactivation is typically carried out by adjusting the pH to
below 4.5 at 95°C, before lowering the temperature to 60°C for
saccharification.

The cause of this negative effect on dextrose yield is not
10 fully understood, but it is assumed that the liquefying α -amylase
preparation employed (e.g. a Termamyl™ product, such as
Termamyl™ 120 L) generates "limit dextrans" (which are poor
substrates for *B. acidopullulyticus* pullulanase) by hydrolysing
1,4- α -glucosidic linkages close to and on both sides of the
15 branching points in amylopectin. Hydrolysis of these limit
dextrans by glucoamylase leads to a build-up of the trisaccharide
panose, which is only slowly hydrolysed by glucoamylase.

The development of a thermostable α -amylolytic enzyme which
does not suffer from this disadvantage would be a significant
20 process improvement, as no separate inactivation step would be
required.

One object of the present invention is to achieve improved
performance of α -amylolytic enzymes in relation to starch
liquefaction processes - e.g. by achieving one or more of the
25 above-outlined improvements - by changing the affinity of the
enzyme for the starch substrate, whereby the modified enzyme
comes into more intimate contact with the substrate.

SUMMARY OF THE INVENTION

30 One aspect of the invention relates to an improved enzymatic
process for liquefying starch employing a modified form of a
liquefying α -amylase, wherein the α -amylase in question is linked
to an amino acid sequence comprising a carbohydrate-binding
domain (*vide infra*).

35 The invention also relates to an improved enzymatic process
for liquefying starch which besides a modified α -amylase also is

treated with a debranching enzyme. The debranching enzyme may be modified by linkage to an amino acid sequence comprising a carbohydrate-binding domain.

Similarly, and also within the scope of the invention, it is envisaged that the use of an analogously modified (i.e. CBD-derivatized) form of a debranching enzyme, such as an isoamylase or a pullulanase, for debranching amylopectin-derived starch fragments (e.g. in connection with the above-outlined saccharification stage of a starch conversion process) will result in enhanced debranching performance, and thereby dextrose yield improvement, in the saccharification procedure.

DETAILED DESCRIPTION OF THE INVENTION

In a first aspect the present invention thus relates to a method for liquefying starch, wherein a starch substrate is treated in aqueous medium with a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of an α -amylase linked (i.e. covalently bound) to an amino acid sequence comprising a carbohydrate-binding domain (CBD).

The invention also relates to an improved enzymatic process for liquefying starch which besides a modified α -amylase also is treated with a debranching enzyme. The debranching enzyme may be modified by linkage to an amino acid sequence comprising a carbohydrate-binding domain.

A further aspect of the present invention relates to a method for saccharifying starch which has been subjected to a liquefaction process, wherein the reaction mixture after liquefaction is treated with a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of an amylopectin-debranching enzyme (e.g. an isoamylase or a pullulanase) linked (i.e. covalently bound) to an amino acid sequence comprising a carbohydrate-binding domain (CBD).

It is to be understood that starch liquefaction processes as referred to in the context of the present invention do not embrace, for example, textile de-sizing processes wherein starch ("size") present in fabrics or textiles (normally cellulosic or cellulose-containing fabrics or textiles) is removed from the

fabric or textile by an enzymatic process.

Carbohydrate-binding domains

A carbohydrate-binding domain (CBD) is a polypeptide amino
5 acid sequence which binds preferentially to a poly- or
oligosaccharide (carbohydrate), frequently - but not
necessarily exclusively - to a water-insoluble (including
crystalline) form thereof.

Although a number of types of CBDs have been described in
10 the patent and scientific literature, the majority thereof -
many of which derive from cellulolytic enzymes (cellulases) -
are commonly referred to as "cellulose-binding domains"; a
typical cellulose-binding domain will thus be a CBD which
occurs in a cellulase. Likewise, other sub-classes of CBDs
15 would embrace, e.g., chitin-binding domains (CBDs which
typically occur in chitinases), xylan-binding domains (CBDs
which typically occur in xylanases), mannan-binding domains
(CBDs which typically occur in mannanases), starch-binding
domains [CBDs which may occur in certain amylolytic enzymes,
20 such as certain glucoamylases, or in enzymes such as
cyclodextrin glucanotransferases ("CGTases")], and others.

CBDs are found as integral parts of large polypeptides or
proteins consisting of two or more polypeptide amino acid
sequence regions, especially in hydrolytic enzymes (hydrolases)
25 which typically comprise a catalytic domain containing the
active site for substrate hydrolysis and a carbohydrate-binding
domain (CBD) for binding to the carbohydrate substrate in
question. Such enzymes can comprise more than one catalytic
domain and one, two or three CBDs, and optionally further
30 comprise one or more polypeptide amino acid sequence regions
linking the CBD(s) with the catalytic domain(s), a region of
the latter type usually being denoted a "linker". Examples of
hydrolytic enzymes comprising a CBD - some of which have
already been mentioned above - are cellulases, xylanases,
35 mannanases, arabinofuranosidases, acetylesterases and
chitinases. CBDs have also been found in algae, e.g. in the red
alga *Porphyra purpurea* in the form of a non-hydrolytic

polysaccharide-binding protein [see P. Tomme et al. Cellulose-Binding Domains - Classification and Properties in Enzymatic Degradation of Insoluble Carbohydrates, John N. Saddler and Michael H. Penner (Eds.), ACS Symposium Series, No. 618
5 (1996)]. However, most of the known CBDs [which are classified and referred to by P. Tomme et al. (*op cit.*) as "cellulose-binding domains"] derive from cellulases and xylanases.

In the present context, the term "cellulose-binding domain" is intended to be understood in the same manner as in the
10 latter reference (P. Tomme et al., *op. cit.*), and the abbreviation "CBD" as employed herein will thus often be interpretable either in the broader sense (carbohydrate-binding domain) or in the - in principle - narrower sense (cellulose-binding domain). The P. Tomme et al. reference classifies more
15 than 120 "cellulose-binding domains" into 10 families (I-X) which may have different functions or roles in connection with the mechanism of substrate binding. However, it is anticipated that new family representatives and additional CBD families will appear in the future.

20 In proteins/polypeptides in which CBDs occur (e.g. enzymes, typically hydrolytic enzymes), a CBD may be located at the N or C terminus or at an internal position.

That part of a polypeptide or protein (e.g. hydrolytic enzyme) which constitutes a CBD *per se* typically consists of
25 more than about 30 and less than about 250 amino acid residues. For example: those CBDs listed and classified in Family I in accordance with P. Tomme et al. (*op. cit.*) consist of 33-37 amino acid residues, those listed and classified in Family IIa consist of 95-108 amino acid residues, those listed and
30 classified in Family VI consist of 85-92 amino acid residues, whilst one CBD (derived from a cellulase from *Clostridium thermocellum*) listed and classified in Family VII consists of 240 amino acid residues. Accordingly, the molecular weight of an amino acid sequence constituting a CBD *per se* will typically
35 be in the range of from about 4kD to about 40kD, and usually below about 35kD.

Enzyme hybrids

Enzyme classification numbers (EC numbers) referred to in the present specification with claims are in accordance with the Recommendations (1992) of the Nomenclature Committee of the
5 International Union of Biochemistry and Molecular Biology, Academic Press Inc., 1992.

As already indicated to some extent (*vide supra*), modified enzymes as referred to herein (in the following also denoted "enzyme hybrids") include species comprising an amino acid
10 sequence of an amylolytic enzyme [which in the context of the present invention may, e.g., be an α -amylase (EC 3.2.1.1), an isoamylase (EC 3.2.1.68) or a pullulanase (EC 3.2.1.41)] linked (i.e. covalently bound) to an amino acid sequence comprising a CBD.

15 Other CBD-containing enzyme hybrids of interest in relation to degradation of starch include, e.g., hybrids comprising an amino acid sequence of a glucan 1,4- α -maltohydrolase (EC 3.2.1.133), a β -amylase (EC 3.2.1.2), a glucoamylase (EC 3.2.1.3), or a neopullulanase (EC 3.2.1.135).

20 CBD-containing enzyme hybrids, as well as detailed descriptions of the preparation and purification thereof, are known in the art [see, e.g., WO 90/00609, WO 94/24158 and WO 95/16782, as well as Greenwood et al., Biotechnology and Bioengineering 44 (1994) pp. 1295-1305]. They may, e.g., be
25 prepared by transforming into a host cell a DNA construct comprising at least a fragment of DNA encoding the cellulose-binding domain ligated, with or without a linker, to a DNA sequence encoding the enzyme of interest, and growing the transformed host cell to express the fused gene. The resulting
30 recombinant product (enzyme hybrid) - often referred to in the art as a "fusion protein" - may be described by the following general formula:

A-CBD-MR-X

35

In the latter formula, A-CBD is the N-terminal or the C-terminal region of an amino acid sequence comprising at least the

carbohydrate-binding domain (CBD) *per se*. MR is the middle region (the "linker"), and X is the sequence of amino acid residues of a polypeptide encoded by a DNA sequence encoding the enzyme (or other protein) to which the CBD is to be linked.

5 The moiety A may either be absent (such that A-CBD is a CBD *per se*, i.e. comprises no amino acid residues other than those constituting the CBD) or may be a sequence of one or more amino acid residues (functioning as a terminal extension of the CBD *per se*). The linker (MR) may be a bond, or a short linking group
10 comprising from about 2 to about 100 carbon atoms, in particular of from 2 to 40 carbon atoms. However, MR is preferably a sequence of from about 2 to about 100 amino acid residues, more preferably of from 2 to 40 amino acid residues, such as from 2 to 15 amino acid residues.

15 The moiety X may constitute either the N-terminal or the C-terminal region of the overall enzyme hybrid.

It will thus be apparent from the above that the CBD in an enzyme hybrid of the type in question may be positioned C-terminally, N-terminally or internally in the enzyme hybrid.

20

Cellulases (cellulase genes) useful for preparation of CBDs

Techniques suitable for isolating a cellulase gene are well known in the art. In the present context, the term "cellulase" refers to an enzyme which catalyses the degradation of cellulose
25 to glucose, cellobiose, triose and/or other cello-oligosaccharides.

Preferred cellulases (i.e. cellulases comprising preferred CBDs) in the present context are microbial cellulases, particularly bacterial or fungal cellulases. Endoglucanases (EC
30 3.2.1.4), particularly mono-component (recombinant) endoglucanases, are a preferred class of cellulases,.

Useful examples of bacterial cellulases are cellulases derived from or producible by bacteria from the group consisting of *Pseudomonas*, *Bacillus*, *Cellulomonas*, *Clostridium*, *Microspora*,
35 *Thermotoga*, *Caldocellum* and Actinomycets such as *Streptomyces*, *Termomonospora* and *Acidothamus*, in particular from the group consisting of *Pseudomonas cellulolyticus*, *Bacillus lautus*,

Bacillus agaradherens, *Cellulomonas fimi*, *Clostridium thermocellum*, *Clostridium stercorearium*, *Microspora bispora*, *Termomonospora fusca*, *Termomonospora cellulolyticum* and *Acidothamus cellulolyticus*.

- 5 The cellulase may be an acid, a neutral or an alkaline cellulase, i.e. exhibiting maximum cellulolytic activity in the acid, neutral or alkaline range, respectively.

A useful cellulase is an acid cellulase, preferably a fungal acid cellulase, which is derived from or producible by fungi from
10 the group of genera consisting of *Trichoderma*, *Myrothecium*, *Aspergillus*, *Phanaerochaete*, *Neurospora*, *Neocallimastix* and *Botrytis*.

A preferred useful acid cellulase is one derived from or producible by fungi from the group of species consisting of *Tri-*
15 *choderma viride*, *Trichoderma reesei*, *Trichoderma longibrachiatum*, *Myrothecium verrucaria*, *Aspergillus niger*, *Aspergillus oryzae*, *Phanaerochaete chrysosporium*, *Neurospora crassa*, *Neocallimastix partriciarum* and *Botrytis cinerea*.

Another useful cellulase is a neutral or alkaline cellulase,
20 preferably a fungal neutral or alkaline cellulase, which is derived from or producible by fungi from the group of genera consisting of *Aspergillus*, *Penicillium*, *Myceliophthora*, *Humicola*, *Irpex*, *Fusarium*, *Stachybotrys*, *Scopulariopsis*, *Chaetomium*, *Mycogone*, *Verticillium*, *Myrothecium*, *Papulospora*, *Gliocladium*, *Cepha-*
25 *losporium* and *Acremonium*.

A preferred alkaline cellulase is one derived from or producible by fungi from the group of species consisting of *Humicola insolens*, *Fusarium oxysporum*, *Myceliophthora thermophila*, *Penicillium janthinellum* and *Cephalosporium* sp., preferably from
30 the group of species consisting of *Humicola insolens* DSM 1800, *Fusarium oxysporum* DSM 2672, *Myceliophthora thermophila* CBS 117.65, and *Cephalosporium* sp. RYM-202.

A preferred cellulase is an alkaline endoglucanase which is immunologically reactive with an antibody raised against a highly
35 purified ~43kD endoglucanase derived from *Humicola insolens* DSM 1800, or which is a derivative of the latter ~43kD endoglucanase and exhibits cellulase activity.

Other examples of useful cellulases are variants of parent cellulases of fungal or bacterial origin, e.g. a parent cellulase derivable from a strain of a species within one of the fungal genera *Humicola*, *Trichoderma* or *Fusarium*.

5

Other proteins (protein genes) useful for preparation of CBDs

Examples of other types of hydrolytic enzymes which comprise a CBD are, as already mentioned, xylanases, mannanases, arabinofuranosidases, acetylerases and
10 chitinases. As also mentioned previously, CBDs have also been found, for example, in certain algae, e.g. in the red alga *Porphyra purpurea* in the form of a non-hydrolytic polysaccharide-binding protein. Reference may be made to P. Tomme et al. (*op cit.*) for further details concerning sources
15 (organism genera and species) of such CBDs. Further CBDs of interest in relation to the present invention include CBDs deriving from glucoamylases (EC 3.2.1.3) or from CGTases (EC 2.4.1.19).

CBDs deriving from such sources will also be generally be
20 suitable for use in the context of the invention. In this connection, techniques suitable for isolating, e.g., xylanase genes, mannanase genes, arabinofuranosidase genes, acetylerase genes, chitinase genes (and other relevant genes) are well known in the art.

25

Isolation of a CBD

In order to isolate a cellulose-binding domain of, e.g., a cellulase, several genetic engineering approaches may be used. One method uses restriction enzymes to remove a portion of the
30 gene and then to fuse the remaining gene-vector fragment in frame to obtain a mutated gene that encodes a protein truncated for a particular gene fragment. Another method involves the use of exonucleases such as *Bal31* to systematically delete nucleotides either externally from the 5' and the 3' ends of the DNA or
35 internally from a restricted gap within the gene. These gene-deletion methods result in a mutated gene encoding a shortened gene molecule whose expression product may then be evaluated for

substrate-binding (e.g. cellulose-binding) ability. Appropriate substrates for evaluating the binding ability include cellulosic materials such as Avicel™ and cotton fibres. Other methods include the use of a selective or specific protease capable of
5 cleaving a CBD, e.g. a terminal CBD, from the remainder of the polypeptide chain of the protein in question

As already indicated (*vide supra*), once a nucleotide sequence encoding the substrate-binding (carbohydrate-binding) region has been identified, either as cDNA or chromosomal DNA, it may then
10 be manipulated in a variety of ways to fuse it to a DNA sequence encoding the enzyme of interest. The DNA fragment encoding the carbohydrate-binding amino acid sequence, and the DNA encoding the enzyme of interest are then ligated with or without a linker. The resulting ligated DNA may then be manipulated in a variety of
15 ways to achieve expression. Preferred microbial expression hosts include certain *Aspergillus* species (e.g. *A. niger* or *A. oryzae*), *Bacillus* species, and organisms such as *Escherichia coli* or *Saccharomyces cerevisiae*.

20 Amylolytic enzymes

Amylases (in particular α -amylases) which are appropriate as the basis for CBD/amylase hybrids of the types employed in the context of the present invention include those of bacterial or fungal origin. Chemically or genetically modified mutants of such
25 amylases are included in this connection. Relevant α -amylases include, for example, α -amylases obtainable from *Bacillus* species, in particular a special strain of *B. licheniformis*, described in more detail in GB 1296839. Relevant commercially available amylases include Duramyl™, Termamyl™, Fungamyl™ and
30 BAN™ (all available from Novo Nordisk A/S, Bagsvaerd, Denmark), and Rapidase™ and Maxamyl P™ (available from Gist-Brocades, Holland), and Optitherm™ (available from Solvay), and Spezym AA™ and Spezyme Delta AA₂ (available from Genencor), and Keistase™ (available from Daiwa).

35 Other amylases (in particular α -amylases) which are appropriate as the basis for CBD/amylase hybrids of the types

employed in the context of the present invention include a hybrid α -amylase consisting of 1-35 N-terminal amino acids of BAN[®] (available from Novo Nordisk) and the C-terminal 36-483 C-terminal amino acids of Termamyl[®] (available from Novo Nordisk) with one or more of the following mutations H156Y, A181T, N190F A209V, Q264S; Termamyl[®] with one or more of the following mutations I201E, D207H, E211Q, H205S; or Maxamyl[™] (available from Gist-brocades/Genencor), with one or more of the following mutations H133Y, N188P,S.

10

Starch- or starch-fragment-debranching enzymes

Isoamylases: isoamylases (EC 3.2.1.68) appropriate as the basis for CBD/isoamylase hybrids of the types employed in the context of the present invention include those of bacterial origin. Chemically or genetically modified mutants of such isoamylases are included in this connection. Relevant isoamylases include, for example, isoamylases obtainable from *Pseudomonas* species, (e.g. *Pseudomonas* sp. SMP1 or *P. amyloclavata* SB15), *Bacillus* species (e.g. *B. amyloliquefaciens*), *Flavobacterium* species or *Cytophaga* (*Lysobacter*) species.

Pullulanases: pullulanases (EC 3.2.1.41) appropriate as the basis for CBD/pullulanase hybrids of the types employed in the context of the present invention include those of bacterial origin. Chemically or genetically modified mutants of such pullulanases are included in this connection. Relevant pullulanases include, for example, pullulanases obtainable from *Bacillus* species (e.g. *B. acidopullulyticus*; such a Promozyme[™], from Novo Nordisk A/S).

30 Plasmids

Preparation of plasmids capable of expressing fusion proteins having the amino acid sequences derived from fragments of more than one polypeptide are well known in the art (see, e.g. WO 90/00609 and WO 95/16782). The expression cassette may be included within a replication system for episomal maintenance in an appropriate cellular host or may be provided without a replication system, where it may become integrated into the host

genome. The DNA may be introduced into the host in accordance with known techniques such as transformation, microinjection or the like.

Once the fused gene has been introduced into the appropriate host, the host may be grown to express the fused gene. Normally it is desirable additionally to add a signal sequence which provides for secretion of the fused gene. Typical examples of useful fused genes are:

- 10 Signal sequence -- (pro-peptide) -- carbohydrate-binding domain -
- linker -- enzyme of interest, or

Signal sequence -- (pro-peptide) -- enzyme of interest -- linker
-- carbohydrate-binding domain,

15

in which the pro-peptide sequence normally contains 5-25 amino acid residues.

The recombinant product may be glycosylated or non-
20 glycosylated.

Determination of α -amylolytic activity (KNU)

The α -amylolytic activity of an enzyme or enzyme hybrid may be determined using potato starch as substrate. This method is
25 based on the break-down (hydrolysis) of modified potato starch, and the reaction is followed by mixing samples of the starch/enzyme or starch/hybrid enzyme solution with an iodine solution. Initially, a blackish-blue colour is formed, but during the break-down of the starch the blue colour becomes weaker and
30 gradually turns to a reddish-brown. The resulting colour is compared with coloured glass calibration standards.

One Kilo Novo α -Amylase Unit (KNU) is defined as the amount of enzyme (enzyme hybrid) which, under standard conditions (i.e. at $37 \pm 0.05^\circ\text{C}$, 0.0003 M Ca^{2+} , pH 5.6) dextrinizes 5.26 g starch
35 dry substance (Merck Amylum solubile).

Test conditions suitable for evaluating the performance of CBD-containing enzyme hybrids in starch processing

Test conditions (e.g. conditions of pH, temperature, calcium concentration etc.) suitable for testing, e.g., CBD/ α -amylase, 5 CBD/isoamylase or CBD/pullulanase enzyme hybrids as described herein will suitably be conditions as already described above in connection with industrial starch conversion processes. Assay methods suitable for determining enzymatic activity under various conditions (e.g. pH, temperature, calcium concentration etc., 10 depending on the nature of the enzyme hybrid) are well known in the art for numerous types of enzymes which are appropriate for linkage to a CBD as described herein, and a person of ordinary skill in the art will readily be able to select assay procedures suitable for evaluating the enzymatic performance of enzyme 15 hybrids as employed in the present context.

The invention also relates to an isolated DNA sequence encoding a hybrid enzyme with amylolytic activity comprising:

- (a) a DNA sequence encoding an amylolytic activity;
- (b) a DNA sequences encoding a CBD; and
- 20 (c) a DNA sequence or fragments thereof encoding the linker sequence shown in SEQ ID no. 21.

It is often a problem of hybrid enzyme comprising an enzyme and a CDB connected via a linker that they are not very stable due to the linker. The inventors have found that when using the 25 linker shown in SEQ ID NO. 21 or essential parts thereof the hybrids are very stable.

The isolated DNA sequence of the invention typically encodes an enzyme with amylolytic activity, such as α -amylase activity, in particular a *Bacillus* α -amylase activity, 30 especially the activity of Termamyl[®] or a variant thereof, or one of the amylolytic activities mentioned above in the section "Amylolytic enzymes". The CBD may be any CBD e.g. the CBDs described above in the section "Carbohydrate-binding domains". In a preferred embodiment the CBD is the CBD of the *Bacillus* 35 *agaradherens* NCIMB No. 40482 alkaline cellulase Cel5A or the CBD-dimer of *Clostridium stercorarium* (NCIMB 11754) XynA..

In a specific embodiment of the invention the isolated DNA sequence is the Termamyl~~III~~-linker-Cel5A-CBD encoded by plasmid pMB492 shown in SEQ ID No. 19.

In a further aspect the invention relates to a DNA
5 construct comprising the isolated DNA sequence of the invention operably linked to one or more control sequences capable of directing the expression of the DNA sequence in a suitable expression host.

The promoter may be any DNA sequence which shows
10 transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell. Examples of suitable promoters for directing the transcription of the DNA encoding the cellulytic enzyme of the invention in bacterial host cells
15 include the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the *Bacillus licheniformis* alpha-amylase gene, the *Bacillus amyloliquefaciens* BAN amylase gene, the *Bacillus subtilis* alkaline protease gene, or the *Bacillus pumilus* xylanase or xylosidase gene, the phage Lambda P_R or P_L
20 promoters, or the *E. coli* lac, trp or tac promoters.

Examples of suitable promoters for use in yeast host cells include promoters from yeast glycolytic genes (Hitzeman et al. (1980) J. Biol. Chem. 255:12073-12080; Alber and Kawasaki (1982) J. Mol. Appl. Gen. 1:419-434) or alcohol dehydrogenase
25 genes (Young et al. (1982) in Genetic Engineering of Microorganisms for Chemicals (Hollaender et al, eds.), Plenum Press, New York), or the TPI1 (US 4,599,311) or ADH2-4c (Russell et al. (1983) Nature 304:652-654) promoters.

To direct the CBD/enzyme hybrid into the secretory pathway
30 of the host cells, a secretory signal sequence (also known as a leader sequence, prepro sequence or pre sequence) may be provided in the expression vector. The secretory signal sequence is joined to the DNA sequence encoding the enzyme hybrid in the correct reading frame. Secretory signal sequences
35 are commonly positioned 5' to the DNA sequence encoding the amylolytic enzyme. The secretory signal sequence may be that normally associated with the amylolytic enzyme or may be from a

gene encoding another secreted protein.

In a preferred embodiment, the expression vector of the invention may comprise a secretory signal sequence substantially identical to the secretory signal encoding
5 sequence of the *Bacillus licheniformis* α -amylase gene, e.g. as described in WO 86/05812.

Also, measures for amplification of the expression may be taken, e.g. by tandem amplification techniques, involving single or double crossing-over, or by multicopy techniques,
10 e.g. as described in US 4,959,316 or WO 91/09129. Alternatively the expression vector may include a temperature sensitive origin of replication, e.g. as described in EP 283,075.

Procedures for ligating DNA sequences encoding the cellulytic enzyme, the promoter and optionally the terminator
15 and/or secretory signal sequence, respectively, and to insert them into suitable vectors containing the information necessary for replication, are well known to persons skilled in the art (cf., for example, Sambrook et al. (1989) supra).

The invention also relates to a recombinant expression
20 vector comprising the DNA construct of the invention, a promoter, and transcriptional and translational stop signals.

It is also an object of the invention to provide a host cell comprising the DNA construct of the invention.

The host cell of the invention, into which the DNA
25 construct or the recombinant expression vector of the invention is to be introduced, may be any cell which is capable of producing the amylolytic enzyme and includes bacteria, yeast, fungi and higher eukaryotic cells.

Examples of bacterial host cells which, on cultivation, are
30 capable of producing the cellulytic enzyme of the invention are grampositive bacteria such as strains of *Bacillus*, in particular a strain of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B. coagulans*, *B. circulans*, *B. lautus*, *B.*
35 *megatherium*, *B. pumilus*, *B. thuringiensis* or *B. agaradherens*, or strains of *Streptomyces*, in particular a strain of *S. lividans* or *S. murinus*, or gramnegative bacteria such as

Echerichia coli. The transformation of the bacteria may be effected by protoplast transformation or by using competent cells in a manner known *per se* (cf. *Sambrook et al.* (1989) *supra*).

5 When expressing the CBD/enzyme hybrid in bacteria such as *E. coli*, the enzyme may be retained in the cytoplasm, typically as insoluble granules (known as inclusion bodies), or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed and the
10 granules are recovered and denatured after which the cellulytic enzyme is refolded by diluting the denaturing agent. In the latter case, the hybrid enzyme may be recovered from the periplasmic space by disrupting the cells, e.g. by sonication or osmotic shock, to release the contents of the periplasmic
15 space and recovering the hybrid enzyme.

The transformed or transfected host cell described above is then cultured in a suitable nutrient medium under conditions permitting the expression of the cellulytic enzyme, after which the resulting cellulytic enzyme is recovered from the culture.

20 The medium used to culture the cells may be any conventional medium suitable for growing the host cells, such as minimal or complex media containing appropriate supplements.

Suitable media are available from commercial suppliers or may be prepared according to published recipes (e.g., in catalogues
25 of the American Type Culture Collection). The cellulytic enzyme produced by the cells may then be recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, precipitating the proteinaceous components of the supernatant
30 or filtrate by means of a salt, e.g., ammonium sulphate, purification by a variety of chromatographic procedures, e.g., ion exchange chromatography, gelfiltration chromatography, affinity chromatography, or the like, dependent on the type of cellulytic enzyme in question.

35 The present invention also relates to methods for producing a CBD/enzyme hybrid of the present invention comprising (a) cultivating a *Bacillus* strain to produce a supernatant

comprising the polypeptide; and (b) recovering the polypeptide.

The present invention also relates to methods for producing a hybrid enzyme of the present invention comprising (a) cultivating a host cell under conditions conducive to
5 expression of the polypeptide; and (b) recovering the polypeptide.

In both methods, the cells are cultivated in a nutrient medium suitable for production of the hybrid enzyme using methods known in the art. For example, the cell may be
10 cultivated by shake flask cultivation, small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the polypeptide to be expressed and/or isolated. The
15 cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art (see, e.g., references for bacteria and yeast; Bennett, J.W. and LaSure, L., eds. (1991) More Gene Manipulations in Fungi, Academic Press, CA).
20 Suitable media are available from commercial suppliers or may be prepared according to published compositions (e.g., in catalogues of the American Type Culture Collection). If the polypeptide is secreted into the nutrient medium, the polypeptide can be recovered directly from the medium. If the
25 polypeptide is not secreted, it is recovered from cell lysates.

The hybrid enzyme may be detected using methods known in the art that are specific for the hybrid enzymes. These detection methods may include use of specific antibodies, formation of an enzyme product, or disappearance of an enzyme
30 substrate. For example, an enzyme assay may be used to determine the activity of the enzyme. Procedures for determining amylolytic activity are known in the art and are described below.

The resulting hybrid enzyme may be recovered by methods
35 known in the art. For example, the hybrid enzyme may be recovered from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration,

extraction, spray-drying, evaporation, or precipitation. The recovered hybrid enzyme may then be further purified by a variety of chromatographic procedures, e.g., ion exchange chromatography, gel filtration chromatography, affinity
5 chromatography, or the like.

The hybrid enzyme of the present invention may be purified by a variety of procedures known in the art including, but not limited to, chromatography (e.g., ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion),
10 electrophoretic procedures (e.g., preparative isoelectric focusing (IEF), differential solubility (e.g., ammonium sulfate precipitation), or extraction (see, e.g., Protein Purification (Janson and Ryden, eds.), VCH Publishers, New York, 1989).

In a final aspect the invention relates to an isolated and
15 purified CBD/enzyme hybrid encoded by the isolated DNA sequence of the invention, in particular the hybrid shown in SEQ ID No. 20.

MATERIALS AND METHODS

20 Materials:

Enzymes and enzyme hybrids:

Termamyl₈-linker-CBD_{EGV} : Hybrid of Termamyl₈ and the fungal CBD_{EGV} from *Humicola insolens* EGV. The construction of the hybrid is described in Example 9.

25

CBD_{CenA}-Termamyl₈ : Hybrid of the CBD_{CenA} from *Cellulomonas fimi* endoglucanase A (CenA) and Termamyl₈ via a linker. The construction of the hybrid is described in Example 8.

30 Termamyl₈ (available from Novo Nordisk A/S)

Plasmids:

pDN1528 (S.Jørgensen et al. (1991) Journal of Bacteriology, vol. 173, No., p-559-567.)

35

pBluescriptKSII- (Stratagene, USA).

pDN1981 (P.L. Jørgensen, C.K.Hansen, G.B.Poulsen and B.Diderichsen (1990) In vivo genetic engineering: homologues recombination as a tool for plasmid construction, Gene, 96, p37-41.)

5

pSJ1678: Described in WO 94/19454; pDN1981: Described by Jørgensen et al. (1990) Gene 96:37-41).

Strains:

- 10 *Bacillus* AC13 NCIMB 40482 (identical to *Bacillus agaradherens* DSM 8721) expressing the endoglucanase enzyme encoding DNA sequence of SEQ ID NO:1.described in Example 1 below

- E. coli* strain: Cells of *E. coli* SJ2 (Diderichsen et al. (1990)
15 J. Bacteriol. 172:4315-4321), which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis* were prepared for and transformed by electroporation using a Gene Pulser™ electroporator from BIO-RAD as described by the supplier.

- 20 *B.subtilis* PL2306 was used as the transformation host strain. It is a cellulase-negative strain developed by introducing a disruption in the transcriptional unit of the known *Bacillus subtilis* cellulase gene in *B.subtilis* strain DN1885(Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B.
25 R., Sjøholm, C. (1990) Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. J. Bacteriol. 172:4315-4321). Not only was the cellulase gene of DN1885 disrupted but also two protease encoding genes where disrupted, namely the *aprE* (Stahl,M.L. and E.Ferrari 1984
30 Replacement of the *Bacillus subtilis* subtilisin structural gene with an In vitro-derived deletion mutation. J.Bacteriol. 158:411-418) and *nprE* (Yang, M.Y. et al 1984 Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation.
35 J.Bacteriol. 160:16-21) genes

The disruption was performed essentially as described in *Bacillus subtilis* and other Gram-Positive Bacteria; A.L.

Sonenshein, J.A. Hoch and Richard Losick, Eds. American Society for Microbiology, 1993, p.618).

Bacillus subtilis: ToC46 (Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of
5 aldB, which encodes alpha-acetolactate decarboxylase, an
exoenzyme from *Bacillus brevis*. J. Bacteriol., 172, 4315-4321)
Was used as a secondary expression host, competent cells and
transformation was performed as described above.

10 Solutions/Media/Reagents

Waxy maize from Cerestar

Corn Starch Cerestar (89% DS) GL 03406 Batch 624362

15 TY and LB agar (as described in Ausubel, F. M. et al. (eds.)
"Current protocols in Molecular Biology". John Wiley and Sons,
1995).

SB: 32 g Tryptone, 20 g Yeast Extract, 5 g NaCl and 5 ml 1 N
20 NaOH are mixed in sterile water to a final volume of 1 liter.
The solution is sterilised by autoclaving for 20 min at 121°C.

10% Avicel: 100 g of Avicel (FLUKA, Switzerland) is mixed with
sterile water to a final volume of 1 litre, and the 10% Avicel
25 is sterilised by autoclaving for 20 min at 121°C.

Buffer: 0.05 M potassium phosphate, pH 7.5

Methods

30 DE determination

DE (dextrose equivalent is defined as the amount of reducing
carbohydrate (measured as dextrose-equivalents) in a sample
expressed as w/w% of the total amount of dissolved dry matter).
It is measured by the neocuproine assay (Dygert, Li
35 Floridana(1965) Anal. Biochem. No 368). The principle of the
neocuproine assay is that CuSO_4 is added to the sample, Cu^{++} is
reduced by the reducing sugar and the formed neocuproine

complex is measured at 450 nm.

General molecular biology methods:

DNA manipulations and transformations were performed using
5 standard methods of molecular biology (Sambrook et al. (1989)
Molecular cloning: A laboratory manual, Cold Spring Harbor
lab., Cold Spring Harbor, NY; Ausubel, F. M. et al. (eds.)
"Current protocols in Molecular Biology". John Wiley and Sons,
1995; Harwood, C. R., and Cutting, S. M. (eds.) "Molecular
10 Biological Methods for Bacillus". John Wiley and Sons, 1990).

Enzymes for DNA manipulations were used according to the
specifications of the suppliers.

Cellulytic Activity

15 Cellulytic activity may be measured in cellulase viscosity
units (CEVU), determined at pH 9.0 with carboxymethyl cellulose
(CMC) as substrate.

Cellulase viscosity units are determined relatively to an
enzyme standard (< 1% water, kept in N₂ atmosphere at -20°C;
20 arch standard at -80°C). The standard used, 17-1187, is 4400
CEVU/g under standard incubation conditions, i.e., pH 9.0, Tris
Buffer 0.1 M, CMC Hercules 7 LFD substrate 33.3 g/l, 40.0°C for
30 minutes.

25 α -amylase-Termamyl[®] Activity

See Novo Nordisk analytical method AF 9/6, available on
request.

EXAMPLES

30 The following examples are put forth so as to provide those
of ordinary skill in the art with a complete disclosure and
description of how to make and use various constructs and
perform the various methods of the present invention and are
not intended to limit the scope of what the inventors regard as
35 their invention. Unless indicated otherwise, parts are parts
by weight, temperature is in degrees centigrade, and pressure
is at or near atmospheric pressure. Efforts have been made to

ensure accuracy with respect to numbers used (e.g., length of DNA sequences, molecular weights, amounts, particular components, etc.), but some deviations should be accounted for.

5 EXAMPLE 1

Cloning of *Bacillus agaradherens* Endoglucanase Gene

Genomic DNA Preparation.

The strain NCIMB 40482 (identical to *Bacillus agaradherens* DSM 8721) was propagated in liquid medium as described in WO 94/01532. After 16 hours of incubation at 30°C and 300 rpm, the cells were harvested, and genomic DNA was isolated by the method described by Pitcher et al. (1989) Lett. Appl. Microbiol. 8:151-156).

15 Genomic Library Construction.

Genomic DNA was partially digested with restriction enzyme Sau3A and size-fractionated by electrophoresis on a 0.7 % agarose gel. Fragments of between 2 and 7 kb in size were isolated by electrophoresis onto DEAE-cellulose paper (Dretzen et al. (1981) Anal. Biochem. 112:295-298). Isolated DNA fragments were ligated to BamHI digested, pSJ1678 plasmid DNA.

PCR Amplification.

In order to obtain the endoglucanase gene as ligated to the pSJ1678 vector, the ligation mixture was used as DNA template in a PCR reaction containing 200 mM of each nucleotide (dATP, dCTP, dGTP and dTTP), 2.5 mM MgCl₂, Expand High Fidelity buffer, 2.0 units of Expand High Fidelity PCR system enzyme mix and 300 nM of each of the following primers:

30

Primer 1 (#9555):

5'-TCACAGATCCTC-GCGAATTGGTGCGGCCGCTNGTNG-ARGARCAYGGNC-3' (SEQ ID No. 3).

35

Primer 1 is a degenerated primer designed to match the amino acid sequence (Val-Val-Glu-Glu-His-Gly-Gln) (SEQ ID No. 4) of

the N-terminal amino acid sequence presented in WO94/01532. The last amino acid is only presented by the first nucleotide of the codon namely C. C is the 3'-nucleotide of the primer.

Furthermore, a NotI site is included at the 5'- end for cloning purposes these nucleotides are underlined. Primer 2 (#9029):

5'-CAGAGCAAGAGATTACGCGC-3' (SEQ ID NO:5).

10 Primer 2 corresponds to a sequence present in the pSJ1678 vector.

The PCR cycling was performed in a Hans Landgraf THERMOCYCLER (Hans Landgraf, Germany), following the profile:

15 1 x (120 seconds at 94°C);
10 x (10 seconds at 94°C; 30 seconds at 55°C; 240 seconds at 72°C);
30 x (10 seconds at 94°C; 30 seconds at 55°C; 180 seconds at 72°C; adding 20 seconds to the keep time at 72°C for each
20 new cycle); and
1 x (300 seconds at 72°C).

The PCR product was gel purified by gel eletrophoresis in a 0.7% agarose gel, and the relevant fragment (approx. 1.7 kb) was excised from the gel and purified using QIAquick Gel
25 extraction Kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50 µl of 10mM Tris-HCl, pH 8.5.

This DNA was used as a template for a PCR re-amplification using the same primers, mixture and cycle profile as above.

30 The PCR product was gel purified by gel eletrophoresis in a 0.7% agarose gel, and the relevant fragment was excised from the gel and purified using QIAquick Gel extraction Kit. The purified DNA was eluted in 50 µl of 10 mM Tris-HCl, pH 8.5.

The purified DNA was digested with NotI and HindIII, gel
35 purified as above, and ligated to the vector pBluescriptII KS- (Stratagene, USA), also digested with NotI and HindIII, and the ligation mixture was used to transform *E. coli* SJ2.

Cells were plated on LB agar plates containing ampicillin (200 µg/ml) supplemented with X-gal (5-Bromo-4-chloro-3-indolyl alpha-D-Galactopyranoside, 50 µg/ml).

5 Identification and Characterization of Positive Clones.

The transformed cells were plated on LB agar plates containing ampicillin (200 µg/ml) supplemented with X-gal (5-Bromo-4-chloro-3-indolyl alpha-D-Galactopyranoside, 50 µg/ml), and incubated at 37°C overnight. The next day white colonies
10 were rescued by restreaking these onto fresh LB-ampicillin agar plates and incubated at 37°C overnight. The day after, single colonies of each clone were transferred to liquid LB medium containing ampicillin (200 µg/ml), and incubated overnight at 37°C with shaking at 250 rpm.

15 Plasmids were extracted from the liquid cultures using QIAgen Plasmid Purification mini kit. 5 µl samples of the plasmids are digested with *NotI* and *HindIII*. The digestions were checked by gel electrophoresis on a 0.7 % agarose gel (NuSieve, FMC). The appearance of a DNA fragment of
20 approximately 1.0 kb indicated a positive clone.

Nucleotide Sequencing the Cloned DNA Fragment.

Qiagen purified plasmid DNA was sequenced with the Taq deoxy terminal cycle sequencing kit (Perkin Elmer, USA) and the
25 primer "Reverse" or the primer "Forward":

Reverse: 5'-GTTTTC-CAGTCACGAC-3' (SEQ ID No.6),

Forward: 5'-GCGGATAACAATTCACACAGG-3' (SEQ ID No. 7).

30 The DNA was sequenced using an Applied Biosystems 373A automated sequencer according to the manufacturers instructions. Analysis of the sequence data is performed according to Devereux et al. (1984) Nucleic Acids Res. 12:387-395).

From this sequence new primers could be designed for
35 performing Inverse PCR [cf. McPherson et al. (eds) in PCR-A practical approach; 1991 IRL Press).

Inverse PCR on Genomic DNA of Strain NCIMB 40482.

Genomic DNA was isolated as described above. 2 mg of pure genomic DNA was digested with EcoRI. The EcoRI was heat inactivated at 65°C for 20 minutes, after which a
5 phenol:chloroform extraction of DNA was performed. DNA was finally ethanol precipitated and resuspended in 20 ml TE.

1 ml of EcoRI digested DNA was ligated with T4-DNA ligase in 100 ml reaction mixture containing T4 ligase buffer and 1 Unit T4-DNA ligase (Boehringer Mannheim, Germany). After 18
10 hours of ligation at 14°C, the ligase was heat inactivated at 68°C for 10 minutes. In order to linearize the circulized genomic DNA fragments prior to Inverse PCR, the ligation mixture was supplemented with 10 U of BstEII (a BstEII site was present internally of the DNA sequence obtained above).

15 50 ml of the BstEII digested ligation mixture was used as template in a PCR reaction containing 200 mM of each nucleotide (dATP, dCTP, dGTP and dTTP), 2.5 mM MgCl₂, Expand High Fidelity buffer, 2.0 units of Expand High Fidelity PCR system enzyme mix, and 300 nM of each of the following primers:

20

Primer 3 (#19719): 5'-TGACCCGTACGGTCCGTGGG-3' (SEQ ID No. 8),
and

Primer 4 (#19720): 5'-GGCTCTTGATTTTGTGTCCACC-3' (SEQ ID No.9).

25 The PCR cycling was performed in a Hans Landgraf THERMOCYCLER (Hans Landgraf, Germany), following the profile:

1 x (120 seconds at 94°C);

10 x (10 seconds at 94°C; 30 seconds at 55°C; 240 seconds at 72°C);

30 30 x (10 seconds at 94°C; 30 seconds at 55°C; 180 seconds at 72°C adding 20 seconds to the keep time at 72°C for each new cycle); and

1 x (300 seconds at 72°C).

The PCR product was gel purified by gel eletrophoresis in a
35 0.7% agarose gel, and the relevant fragment (approx. 4-5 kb) was excised from the gel and purified using QIAquick Gel extraction Kit. The purified DNA was eluted in 50 µl of 10mM

Tris-HCl, pH 8.5.

Nucleotide Sequencing the Inverse-PCR DNA Fragment.

Qiagen purified DNA was sequenced with the Taq deoxy
5 terminal cycle sequencing kit (Perkin Elmer, USA), and the
primer 1, 3 and 4 described above, using an Applied Biosystems
373A automated sequencer according to the manufacturers
instructions. Analysis of the sequence data is performed
according to Devereux et al. (1984) *supra*). Based upon the
10 obtained sequence two new primers were designed in order to
clone the alkaline endoglucanase as presented as SEQ ID No. 12.
The primers were #20887 (SEQ ID No. 10) and #100084 (SEQ ID NO.
14) as described below.

15 **EXAMPLE 2**

Expression of the Alkaline Endoglucanase in *Bacillus subtilis*

The nucleotide sequence in SEQ ID No. 12 was cloned by PCR
for introduction in an expression plasmid pDN1981.

PCR was performed as described below on 500 ng of genomic
20 DNA, using the following two primers containing NdeI and KpnI
(the KpnI site is conveniently present in the amplified
sequence) restriction sites for introducing the endoglucanase
encoding DNA sequence to pDN1981 for expression:

25 **Primer 5 (#20887):**

5'-GTA GGC TCA GTC ATA TGT TAC ACA TTG AAA GGG GAG GAG AAT CAT
GAA AAA GAT AAC TAC TAT TTT TGT CG-3' (SEQ ID No. 10), and

30 **Primer 7 (#100084):**

5'- CCT CGC GAG GTA CCA GCG GCC GCG TAC CAC CAA TTA AGT ATG GTA
C -3' (SEQ ID No. 14)

The underlined nucleotides of Primer 5 corresponds to the NdeI
site, and the underlined nucleotides in the Primer 7 is part of
35 the KpnI site present in the sequence.

Using the ExpandTM Long Template PCR system (available from
Boehringer Mannheim, Germany) amplification was performed using

a mixture consisting of (Buffer 1 diluted 10 times) and 200 μ M of each dNTP, 2.5 units of Enzyme mix (Boehringer Mannheim, Germany) and 500 pmol of each primer.

The PCR reactions was performed using a DNA Thermal Cycler (available from Landgraf, Germany). One incubation at 94°C for 2 minutes followed by ten cycles of PCR performed using a cycle profile of denaturation at 94°C for 10 seconds, annealing at 55°C for 30 seconds, and extension at 68°C for 4 minutes. Followed by 25 cycles of PCR performed using a cycle profile of denaturation at 94°C for 10 seconds, annealing at 55°C for 30 seconds, and extension at 68°C for 3 minutes (this duration of extension is extended with 20 seconds for each of the 25 cycles).

Aliquots of 10 μ l of the amplification product is analysed by electrophoresis in 0.7 % agarose gels (NuSieve, FMC) with ReadyLoad 100bp DNA ladder (GibcoBRL, Denmark) as a size marker.

After PCR cycling, the PCR fragment was purified using QIAquick PCR column Kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50 μ l of 10mM Tris-HCl, pH 8.5, digested with NdeI and KpnI, and purified and ligated to digested pDN1981. The ligation mixture was used to transform *B. subtilis* PL2304.

Competent cells were prepared and transformed as described by Yasbin et al. [Yasbin R E, Wilson G A & Young F E; Transformation and transfection in lysogenic strains of *Bacillus subtilis* : evidence for selective induction of prophage in competent cells; J Bacteriol 1975 121 296-304].

30 Isolation and Test of *Bacillus subtilis* Transformants

The transformed cells were plated on LB agar plates containing 10 mg/ml Kanamycin, 0.4% glucose, 10 mM KH₂PO₄ and 0.1% AZCL HE-cellulose (Megazyme, Australia), and incubated at 37 °C for 18 hours. Endoglucanase positive colonies were identified as colonies surrounded by a blue halo.

Each of the positive transformants were inoculated in 10 ml

TY-medium containing 10 mg/ml Kanamycin. After 1 day of incubation at 37°C and stirring at 250 rpm, 50 ml supernatant was removed. The endoglucanase activity was identified by adding 50 ml supernatant to holes punched in the agar of LB agar plates containing 0.1 % AZCL HE-cellulose.

After 16 hours of incubation at 37 °C, blue halos surrounding holes indicated expression of the endoglucanase in *Bacillus subtilis*.

10 EXAMPLE 3

Analysis of the Cloned Sequence.

The protein sequence derived from the cloned endoglucanase gene shows an endoglucanase of the following composition:

Amino acid residues 1 to 26 correspond to a signal peptide; amino acid residues 27 to 326 constitute the actual endoglucanase (homologues to other family 5 glycosyl hydrolases); amino acid residues 327 to 354 correspond to a linker; amino acid residues 355 to 400 correspond to a cellulose binding domain (as described in Example 3); amino acid residues 401 to 416 correspond to a linker; and amino acid residues 417 to 462 constitute a second cellulose binding domain (highly homologues to the first one (at amino acid residues 355 to 400)).

The molar extinction coefficient was determined as 146,370. The molecular weight was approximately 52 kD.

For the protein without the signal sequence the molar extinction coefficient was determined as 146.370. The molecular weight was approximately 49 kD.

The enzyme has no cysteine, and the charged amino acids give a calculated pI of around 4.

EXAMPLE 4

Subcloning of a partial *Termamyl* sequence.

The α -amylase gene encoded on pDN1528 was PCR amplified for introduction of a BamHI site in the 3'-end of the coding region. The PCR and the cloning was done as follows.

Approximately 10 to 20 ng of plasmid pDN1528 was PCR amplified in HiFidelity[®] PCR buffer (Boehringer Mannheim, Germany) supplemented with 200 μ M of each dNTP, 2.6 units of HiFidelity[®] Expand enzyme mix, and 300 pmol of each primer:

5

Primer 8, #5289

5'-GCT TTA CGC CCG ATT GCT GAC GCT G -3' (SEQ ID No. 15)

Primer 9, #26748

10 5'-GCG ATG AGA CGC GCG GCC GCC TAT CTT TGA ACA TAA ATT GAA ACG
GAT CCG -3' (SEQ ID No. 16)

Restriction site BamHI are underlined.

The PCR reactions was performed using a DNA thermal cycler
15 (Landgraf, Germany). One incubation at 94°C for 2 minutes, 30
seconds at 60°C and 45 seconds at 72°C followed by ten cycles
of PCR performed using a cycle profile of denaturation at 94°C
for 30 seconds, annealing at 60°C for 30 seconds, and extension
at 72°C for 45 seconds and twenty cycles of denaturation at
20 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 45
seconds (at this elongation step 20 seconds are added every
cycle). 10 μ l aliquots of the amplification product was
analysed by electrophoresis in 1.0 % agarose gels (NuSieve,
FMC) with ReadyLoad 100bp DNA ladder (GibcoBRL, Denmark) as a
25 size marker.

40 μ l aliquots of the PCR product generated as described
above were purified using QIAquick PCR purification kit
(Qiagen, USA) according to the manufacturer's instructions. The
purified DNA was eluted in 50 μ l of 10mM Tris-HCl, pH 8.5. 25
30 μ l of the purified PCR fragment was digested with BamHI and
PstI, electrophoresed in 1.0% low gelling temperature agarose
(SeaPlaque GTG, FMC) gels, the relevant fragment was excised
from the gel, and purified using QIAquick Gel extraction Kit
(Qiagen, USA) according to the manufacturer's instructions. The
35 isolated DNA fragment was then ligated to BamHI-PstI digested
pBluescriptII KS- and the ligation mixture was used to

transform *E.coli* SJ2.

Cells were plated on LB agar plates containing ampicillin (200 μ g/ml) and supplemented with X-gal (5-Bromo-4-chloro-3-indolyl alpha-D-galactopyranoside, 50 μ g/ml), and incubated at 37°C over night. Next day white colonies were re-streaked onto fresh LB-ampicillin agar plates and incubated at 37°C over night. The next day single colonies were transferred to liquid LB medium containing (200 μ g/ml) and incubated overnight at 37°C with shaking at 250 rpm.

Plasmids were extracted from the liquid cultures using QIAgen Plasmid Purification mini kit (Qiagen, USA) according to the manufacturer's instructions. 5 μ l samples of the plasmids were digested with PstI and BamHI. The digestions were checked by gelelectrophoresis on a 1.0% agarose gel (NuSieve, FMC). One positive clone, containing the PstI-BamHI fragment containing part of the alfa-amylase gene, was designated pMB335. This plasmid was then used in the construction of α -amylase-CBD hybrids.

In vitro amplification of the linker and the most C-terminal CBD of *Bacillus agaradherens* NCIMB No. 40482.

Approximately 100 to 200 ng of chromosomal DNA obtained from *Bacillus agaradherens* NCIMB No. 40482 (as described in the Examples 1 to 3 above) was PCR amplified in HiFidelity[®] PCR buffer (Boehringer Mannheim, Germany) supplemented with 200 μ M of each dNTP, 2.6 units of HiFidelity[®] Expand enzyme mix, and 300 pmol of each primer:

Primer 10, #110150A

5'- GCT GCA GGA TCC GTT TCA ATT TAT GTT CAA AGA TCT GAT CCA GAT TCA GGA G -3' (SEQ ID No. 17)

Primer 11, #100084

5'-CCT CGC GAG GTA CCA GCG GCC GCG TAC CAC CAA TTA AGT ATG GTA C-3' (SEQ ID NO. 18)

Restriction sites BamHI and NotI are underlined.

The primers were designed to amplify the linker and most C-

terminal CBD of the endoglucanase encoding gene of *Bacillus agaradherens* NCIMB No. 40482 described in the Examples above).

The PCR reaction was performed using a DNA thermal cycler (Landgraf, Germany). One incubation at 94°C for 2 minutes, 30 seconds at 60°C and 45 seconds at 72°C followed by ten cycles of PCR performed using a cycle profile of denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 45 seconds and twenty cycles of denaturation at 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds (at this elongation step 20 seconds are added every cycle). 10 µl aliquots of the amplification product was analysed by electrophoresis in 1.5 % agarose gels (NuSieve, FMC) with ReadyLoad 100bp DNA ladder (GibcoBRL, Denmark) as a size marker.

Cloning by polymerase chain reaction (PCR):

Subcloning of PCR fragments.

40 µl aliquots of the PCR products generated as described above were purified using QIAquick PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50 µl of 10mM Tris-HCl, pH 8.5. 25 µl of the purified PCR fragment was digested with NotI and partially digested with BamHI, electrophoresed in 1.5% low gelling temperature agarose (SeaPlaque GTG, FMC) gels, the relevant fragment was excised from the gels, and purified using QIAquick Gel extraction Kit (Qiagen, USA) according to the manufacturer's instructions. The isolated DNA fragment was then ligated to BamHI-NotI digested pMB335 and the ligation mixture was used to transform *E.coli* SJ2.

Identification and characterization of positive clones.

Cells were plated on LB agar plates containing λ (200 µg/ml) and incubated at 37°C over night. Next day colonies were restreaked onto fresh LB-ampicillin agar plates and incubated at 37°C over night. The next day single colonies were

transferred to liquid LB medium containing (200 µg/ml) and incubated overnight at 37°C with shaking at 250 rpm.

Plasmids were extracted from the liquid cultures using QIAGEN Plasmid Purification mini kit (Qiagen, USA) according to the manufacturer's instructions. Five-µl samples of the plasmids were digested with BamHI and NotI. The digestions were checked by gelelectrophoresis on a 1.5% agarose gel (NuSieve, FMC). The appearance of a DNA fragment of the same size as seen from the PCR amplification indicated a positive clone.

One positive clone, containing the fusion construct of the α-amylase gene and the CBD of *Bacillus agaradherens* NCIMB No. 40482 alkaline cellulase Cel5A, was designated MBamYC5ANewlink.

Cloning of the fusion construct into a *Bacillus* based expression vector.

The pDN1528 vector contains the amyL gene of *B. licheniformis* this gene is actively expressed in *B. subtilis* resulting in the production of active α-amylase appearing in the supernatant. For expression purposes the DNA encoding the fusion protein as constructed above was introduced to pDN1528.

This was done by digesting p MBamYC5ANewlink and pDN1528 with SalI-NotI, purifying the fragments and ligating the 4.7 kb pDN1528 SalI-NotI fragment with the 0.5 kb pMBamYC5ANewlink SalI-NotI fragment. This created an inframe fusion of the hybrid construction with the Termamyl gene. See sequence for PMB492 (SEQ ID No. 19).

The ligation mixture was used to transform competent cells of PL2306. Cells were plated on LB agar plates containing chloramphenicol (6 µg/ml), 0.4% glucose and 10mM potassium hydrogen phosphate and incubated at 37°C over night. Next day colonies were restreaked onto fresh LBPG chloramphenicol agar plates and incubated at 37°C over night. The next day single colonies of each clone were transferred to liquid LB medium containing chloramphenicol (6 µg/ml) and incubated overnight at 37°C with shaking at 250 rpm.

Plasmids were extracted from the liquid cultures using QIAGEN Plasmid Purification mini kit (Qiagen, USA) according to

the manufacturer's instructions, however the resuspension buffer was supplemented with 1 mg/ml of Chicken Egg White Lysozyme (SIGMA, USA) prior to lysing the cells at 37°C for 15 minutes. 5 µl samples of the plasmids were digested with BamHI and NotI. The digestions were checked by gelelectrophoresis on a 1.5% agarose gel (NuSieve, FMC). The appearance of a DNA fragment of the same size as seen from the PCR amplification indicated a positive clone. One positive clone was designated MB492.

10

Expression, secretion and functional analysis of the fusion protein.

The clone MB492 (expressing Termamyl fused to *Bacillus agaradherens*-Cel5A-linker-CBD) was incubated for 20 hours in SB-medium at 37°C and 250 rpm. 1 ml of cell-free supernatant was mixed with 200 µl of 10% Avicel. The mixture was left for 1 hour incubation at 0°C. After this binding of CBD to Avicel the Avicel with CBD was spun 5 minutes at 5000g. The pellet was re-suspended in 100 µl of SDS-page buffer, boiled at 95°C for 5 minutes, spun at 5000g for 5 minutes and 25 µl was loaded on a 4-20% Laemmli Tris-Glycine, SDS-PAGE NOVEX gel (Novex, USA). The samples were electrophoresed in a Xcell™ Mini-Cell (NOVEX, USA) as recommended by the manufacturer, all subsequent handling of gels including staining with comassie, destaining and drying were performed as described by the manufacturer.

The appearance of a protein band of approx. 60 kDa, indicated expression in *B.subtilis* of the Termamyl-Linker-CBD fusion encoded on the plasmid pMB492 (SEQ ID No. 19). The expression protein sequence of the fusion construction of pMB492 is shown in SEQ ID No. 20.

The linker region of interest as described in this example is the specific sequence:

SDPDSGEPDPTPPSDPG (SEQ ID No. 21)

35

Example 5

Isolation of genomic DNA from *Clostridium stercorarium* NCIMB 11754.

Clostridium stercorarium NCIMB 11754 was grown anaerobically at 60°C in specified media as recommended by The National
5 Collections of Industrial and Marine Bacteria Ltd. (Scotland). Cells were harvested by centrifugation.

Genomic DNA was isolated as described by Pitcher et al. (Pitcher, D. G., Saunders, N. A., Owen, R. J. (1989). Rapid extraction of bacterial genomic DNA with guanidium thiocyanate.
10 Lett. Appl. Microbiol., 8, 151-156).

In vitro amplification of the CBD-dimer of *Clostridium stercorarium* (NCIMB 11754) XynA.

Approximately 100 to 200 ng of genomic DNA (isolated as
15 described above) was PCR amplified in HiFidelity[®] PCR buffer (Boehringer Mannheim, Germany) supplemented with 200 µM of each dNTP, 2.6 units of HiFidelity[®] Expand enzyme mix, and 300 pmol of each primer:

20 Primer 12, #114135

5'-GCT GCA GGA TCC GTT TCA ATT TAT GTT CAA AGA TCT CCA ACT CCT
GCC CCA TCT CAA AGC-3' (SEQ ID NO. 22)

Primer 13, #110151

25 5'-GCG ATG AGA CGC GCG GCC GCT ACT ACC AGT CAA CAT TAA CAG GAC
CTG AG -3' (SEQ ID NO. 23)

Restriction sites BamHI and NotI are underlined.

The primers were designed to amplify the DNA encoding the
30 Cellulose Binding Domain of the XynA encoding gene of *Clostridium stercorarium* (NCIMB 11754), the DNA sequence was extracted from the database GenBank under the accession number D13325.

The PCR reaction was performed using a DNA thermal cycler
35 (Landgraf, Germany). One incubation at 94°C for 2 minutes, 30 seconds at 60°C and 45 seconds at 72°C followed by ten cycles of PCR performed using a cycle profile of denaturation at 94°C

for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72°C for 45 seconds and twenty cycles of denaturation at 94°C for 30 seconds, 60°C for 30 seconds and 72°C for 45 seconds (at this elongation step 20 seconds are added every 5 cycle). 10 µl aliquots of the amplification product was analyzed by electrophoresis in 1.0 % agarose gels (NuSieve, FMC) with ReadyLoad 100bp DNA ladder (GibcoBRL, Denmark) as a size marker.

10 Cloning by polymerase chain reaction (PCR):

Subcloning of PCR fragments.

40 µl aliquots of the PCR products generated as described above are purified using QIAquick PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. The purified 15 DNA is eluted in 50 µl of 10 mM Tris-HCl, pH 8.5. 25 µl of the purified PCR fragment is digested with BamHI and EagI, electrophoresed in 1.0% low gelling temperature agarose (SeaPlaque GTG, FMC) gels, the relevant fragment is excised from the gels, and purified using QIAquick Gel extraction Kit 20 (Qiagen, USA) according to the manufacturer's instructions. The isolated DNA fragment is then ligated to BamHI-NotI digested pMB335 and the ligation mixture is used to transform *E.coli* SJ2.

The following steps were then performed as described above:

25

-Identification and characterisation of positive clones.

-Cloning of the fusion construct into a *Bacillus* based expression vector.

-Expression, secretion and functional analysis of the 30 fusion protein.

The appearance of a protein band of approximately 87 kDa on the comassie stained SDS-PAGE, shows positive expression of the hybrid in *Bacillus subtilis*.

The resulting hybrid is thus expressed in *Bacillus subtilis* 35 clone MBXynCBD2 and is encoded in the DNA sequence SEQ ID No. 24 which can be translated to the protein sequence shown in SEQ ID No. 25.

EXAMPLE 6**CBD_{Cel15A}-linker-Termamyl₁₀₀ starch processing**

It is investigated whether or not CBD_{Cel15A}-linker-Termamyl₁₀₀
 5 (i.e. *Bacillus agaradherens* NCIMB 40482 endoglucanase C-terminal CBD linked to Termamyl₁₀₀ via the linker shown in SEQ ID No. 21 constructed as described in Example 4) gives an improved liquefaction of starch per µg enzyme protein/g dry substance compared to Termamyl₁₀₀ at pH 6.0 and 40 ppm Ca²⁺.

10 A shaking oil bath is heated to 105°C. Two starch slurries (30% DS with 40 ppm Ca⁺⁺) are prepared, the pH is adjusted to 6.0 with NaOH. CBD_{Cel15A}-linker-Termamyl₁₀₀ and Termamyl₁₀₀, respectively, are well mixed into the slurries.

From each slurry four portions of 10 g each are taken. Each
 15 portion are placed in an Erlenmeyer flask with screw cap. The flasks were placed in the oil bath for 8 minutes at 105°C and then 90 minutes at 95°C.

After 7 minutes and 45 seconds in the oil bath, the thermostat of the oil bath is adjusted to 95.4°C and 2 litre
 20 oil at room temperature are added to the oil bath. A clock is started and samples (1 flask of each slurry) are taken after 20, 40, 60, and 90 minutes. 2 drops of 1 N HCl is added to each flask to inactivate the amylase.

The DE-value is then determined as a function of time to
 25 compare the starch liquefaction per µg enzyme/g DS of CBD_{Cel15A}-linker-Termamyl₁₀₀ with Termamyl₁₀₀.

EXAMPLE 7**Construction of the CBD_{CenA} expression vector pCBDT001.**

30 The gene fragment encoding the 103 residue CBD_{CenA} from *Cellulomonas fimi* endoglucanase A (CenA) was cloned in the high expression vector pTugEO7K3. Appropriate restriction sites were introduced at the 5' and 3' ends of the CBD_{CenA} gene by PCR. Each PCR mixture (50 ml total volume) contained 25 ng template
 35 DNA (pTZ18R-1.6cenA; Damude 1995 Doctoral thesis, University of British Columbia, Canada), 25-50 pmole primers (5'SAENH and 3'SAENH), 10 % dimethyl sulfoxide, 0.4 mM 2'-deoxynucleotide

5'-triphosphates, and 1U Vent DNA polymerase in "Thermopol" buffer (New England BioLabs). Twenty successive cycles of denaturation at 94°C for 30 seconds, followed by annealing at 55°C for 30 seconds, and primer extension at 72°C for 54
5 seconds were performed. A *SpeI* site (underlined) was introduced at the 5' end of the *CBD_{CenA}* gene fragment, using the oligonucleotide (5'SAENH)

Primer 14

10 5'-AGGTCTACTAGTCCCGGCTGCCGCGTCGAC-3' (SEQ ID No. 27)

as primer. *EcoRI* (underlined), *NheI* (*in bold*) and *HindIII* (*in italics*) restriction sites were introduced at the 3' end of the *CBD_{CenA}* sequence using the oligonucleotide (3'SAENH)

15

Primer 15

5 '-CCGATTAAAGCTTATTAGCTAGCACGGAATTCCGTGGGGCTGGTCGTCGGCAC-3'
(SEQ ID No. 28)

20 as primer. The resulting 0.38 kb PCR fragment was digested with *SpeI* and *HindIII* and ligated in frame with the Cex leader peptide at the *NheI*-*HindIII* site of pTugE07K3 , previously cut with *NheI* and *HindIII* to remove the *CBD_{Cex}* gene fragment. The final construct pCBDT001 was verified by restriction and PCR
25 analysis.

2. Construction of the CBD-Termamyl₁ hybrid expression vector pNAMK 1.0 .

The plasmid pSJ3368 a derivative of pDN1528 (S.Jørgensen et
30 al. (1991) Journal of Bacteriology, vol. 173, No., p-559-567.) containing the *Termamyl₁* gene, was isolated from *Bacillus* by standard methods. Appropriate restriction sites for recloning the *Termamyl₁* gene fragment in the *E. coli* vector pCBDT001 and for the construction of the hybrids were introduced by PCR.
35 Each PCR reaction mixture (50 ml total volume) contained 15 ng template DNA (pSJ3368), 3 pmol primers (PAM1 and PAM2), 2 mM MgSO₄, 10 % dimethyl sulfoxide, 0.4 mM 2'-deoxynucleotide 5'-

triphosphates and 1U Vent DNA polymerase in "Thermopol" buffer (New England BioLabs). Thirty successive cycles were performed as follows: denaturation at 95°C for 1 min, annealing at 55°C for 1 min and primer extension at 72°C for 1.54 min.

- 5 A *NheI* (underlined) and *NcoI* site were introduced at the 5' end of the gene with the oligonucleotide (PAM1)

Primer 16

5'-TCATGAGCCATGGCTAGCGCAAATCTTAATGGGACGCTGATG-3'

- 10 (SEQ ID NO. 29)

as primer. An *SpeI* (*in bold*) and *HindIII* site (underlined) were introduced at the 3' end of the Termamyl gene using the oligonucleotide (PAM2)

15

Primer 17

5'-ATGACTA**AGCTTAC** TTACTTAGTGATGGTGATGGTGATGACTAGTTCTTTGAA
CATAAATTGAAACCGA-3' (SEQ ID NO. 30)

- 20 as primer. This also introduced a His₆-tag (*in italics*) for easy purification of the hybrid protein by immobilized metal affinity chromatography (IMAC), and a stop codon immediately preceding the *HindIII* restriction sequence. The resulting 1.5 kb fragment was digested with *NheI* and *HindIII* and cloned in
25 frame with the CBD_{CenA} at the *NheI*-*HindIII* site of pCBDT001 to give pNAMK 1.0. The construct was verified by restriction digesting with *NheI* and *HindIII* and by automated sequencing.

CBD_{CenA}-PTPTTP-Termamyl~~®~~ production and purification

- 30 Overnight cultures of *E. coli* JM101, harboring plasmid pNAM1.0, were diluted 500-fold in terrific broth (TB; 12 g tryptone, 24 g yeast extract, 9.8 g K₂HPO₄, 2.2 g KH₂PO₄ and 8 g (10 ml) glycerol in 1l) (Sambrook et al., 1989) (ref: Sambrook J., Fritsch, E.F., & Maniatis, T. (1989) Molecular cloning: a
35 laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.) supplemented with 1.25 mM CaCl₂ and 100 mg kanamycin per ml and grown at 30°C to an A₆₀₀

of 3.0-5.0. Protein production was induced by the addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM. The cultures were incubated for an additional 18 hours at 30°C by which time the CBD-Termamyl~~8~~ hybrid had leaked into the culture medium. Cells were removed by centrifugation at 4°C for 10 minutes at 13,000 x g. The protein was precipitated from the clarified supernatant with 70% (NH₄)₂SO₄ with stirring overnight at 4°C. Proteins were recovered by centrifugation at 11,000 x g and the pellet was dissolved in 20 mM Tris-HCl, pH 8.0 (binding buffer). After further centrifugation at 15000 x g, the clarified supernatants was loaded onto a Ni²⁺ agarose column (Novagen, Markham, ON). The column was washed with 40 mM imidazole, 200 mM NaCl, 20 mM Tris-HCl, pH 8.0 (wash buffer). Bound proteins were eluted with a gradient of imidazole (0-500 mM) in 20 mM Tris-HCl buffer containing 500 mM NaCl. CaCl₂ was immediately added to the fractions to a final concentration of 1 mM to stabilize the protein. Fractions were analysed on SDS-PAGE (12%) and by activity measurements.

The NAM1.0 nucleotide sequence is shown in SEQ ID NO. 31 and can be translated into the amino acid sequence shown in SEQ ID No. 32.

EXAMPLE 8

Termamyl linker fungal CBD from *Humicola insolens* EGV.

pNAMK6.1 (Termamyl~~8~~-linker-CBDEGV)

The Termamyl vector NAM 2.0 for C-terminal CBD:

Each PCR reaction mixture (50 ml total volume) contained 15 ng template DNA (pSJ3368), 3 pmol primers (5Term2 and 3Term2), 2 mM MgSO₄, 10 % dimethyl sulfoxide, 0.4 mM 2'-deoxynucleotide 5'-triphosphates and 1U Vent DNA polymerase in "Thermopol" buffer (New England BioLabs). Thirty successive cycles were performed as follows: denaturation at 95°C for 1 min, annealing at 55°C for 1 min and primer extension at 72°C for 1.54 min.

NheI (underlined) and *EcoRI* (**in bold**) sites were introduced at the 5' end of the *Termamyl* gene with the oligonucleotide (5Term2)

Primer 18

5'-CATATGGCTAGCGAATTCGCAAATCTTAATGGGACGCTG-3' (SEQ ID NO. 33)

5 as primer. *StuI* (underlined), *SpeI* (*in bold*) and *HindIII* sites (*in italics*) were introduced at the 3' end of the *Termamyl* gene using the oligonucleotide (3Term2)

Primer 19

10 5'-AAGCTTACTAGTAGGCCTTCTTTGAACATAAATT GAAA-3' (SEQ ID NO. 34)

as primer. The construct was verified by restriction digesting and by automated sequencing.

15 The fungal CBD vector:

pCBDT006 was obtained by cloning the gene fragment encoding *CBD_{EGV}* from *Humicola insolens* endoglucanase V (WO 91/17243) in pTugEO7K3. Appropriate restriction sites were introduced at the 5' and 3' ends of the *CBD_{EGV}* gene by PCR. Each PCR mixture (50 ml total volume) contained 25 ng template DNA 25-50 pmole primers (N137 and N1PTcs), 10 % dimethyl sulfoxide, 0.4 mM 2'-deoxynucleotide 5'-triphosphates, and 1U Vent DNA polymerase in "Thermopol" buffer (New England BioLabs). Twenty successive cycles of denaturation at 96°C for 45 seconds, followed by 25 annealing at 50°C for 60 seconds, and primer extension at 72°C for 35 seconds were performed. The last cycle was followed by extension at 72°C for 90 seconds.

NheI (underlined), *EcoRI* (*in bold, underlined*), *StuI* (*in bold*) restriction site were introduced before the artificial 30 linker (*in small letters, italics*), *SpeI* (*in italics, underlined*) and *Eco47III* (*in small, bold*) sites were introduced after the linker at the 3' end of the *CBD_{EGV}* sequence using the oligonucleotide (5CBDT6)

35 Primer 20

5 '- CCATGGGCTAGCCCTGAATTCAGGCCTccaacccccACTAGTCCGagcgctCCC
AGCGGCTGCACTGCTG -3' (SEQ ID No. 35)

as primer. A *Hind*III (underlined) restriction site was introduced at the 3' end of the *CBD_{EGV}* sequence using the oligonucleotide (3CBDT6)

5

Primer 21

5'- AGCCTAAGCTTACAGGCACTGATGGTACCAGT -3' (SEQ ID No. 36)

as primer. The resulting 0.18 kb PCR fragment was digested with
10 *Nhe*I and *Hind*III and ligated in frame with the Cex leader peptide at the *Nhe*I-*Hind*III site of pTugEO7K3 , previously cut with *Nhe*I and *Hind*III to remove the *CBD_{Cex}* gene fragment. The final construct pCBDT006 was verified by restriction and PCR analysis.

15

Construction of the hybrid NAMK6.1 (Termamyl-~~linker~~-*CBD_{EGV}*)

The Termamyl~~linker~~ vector NAM2.0 was digested with *Nhe*I and *Stu*I and the resulting 1.48 kb fragment was gel purified using the Gene Clean (Bio101) kit and ligated in frame with the *CBD_{EGV}*
20 encoding fragment in pCBDT006, previously cut with *Nhe*I and *Stu*I to give pNAMK6.1.

The product has the following characterization MW 60863. Total 537 amino acid residues. First the Termamyl~~linker~~ catalytic amylase then the linker in one letter codes:

25 RPPTPTSPSAPS (SEQ ID No. 37) and finally 38 residues from the fungal CBD. Complete nucleotide Sequence for pNAMK6.1 (pTugK with Termamyl~~linker~~-*CBD_{EGV}* insert) is shown in SEQ ID No. 26.

Example 9

30 **Termamyl~~linker~~-*CBD_{EGV}* starch processing**

It was investigated whether or not the Termamyl~~linker~~-*CBD_{EGV}* (Termamyl~~linker~~ fungal CBD from *Humicola insolens* EGV constructed as described in Example 9 above) gives a better liquefaction of starch per μ g enzyme protein/g dry substance
35 compared to Termamyl~~linker~~ at pH 6.0 and 40 ppm Ca^{2+} .

A shaking oil bath was heated to 105°C. Three starch slurries (30% DS with 40 ppm Ca^{++}) were prepared, the pH was

adjusted to 6.0 with NaOH. The enzyme was well mixed into the slurries according to the scheme:

- Slurry 1: Termamyl[®]-linker-CBDEGV 10.9 µg/g DS starch
 5 Slurry 2: Termamyl[®]-linker-CBDEGV 8.72 µg/g DS starch
 Slurry 3: Termamyl[®] 10.9 µg/g DS starch

From each slurry four portions of 10 g each were taken. Each portion were placed in an Erlenmeyer flask with screw cap. The
 10 flasks were placed in the oil bath for 8 minutes at 105°C and then 90 minutes at 95°C.

After 7 minutes and 45 seconds in the oil bath, the thermostat of the oil bath was adjusted to 95.4°C and 2 litre oil at room temperature were added to the oil bath. A clock was
 15 started and samples (1 flask of each slurry) were taken after 20, 40, 60, and 90 minutes. 2 drops of 1N HCl was added to each flask to inactivate the amylase.

DE-determinations as function of time:

Minutes	Termamyl [®] - linker- CBDEGV 10.9 µg/g DS	Termamyl [®] - linker- CBDEGV 8.72 µg/g DS	Termamyl [®] 10.9 µg/g DS
20	6.1	5.6	5.3
40	9.2	7.4	7.7
60	11.6	10.2	9.1
90	14.6	13.4	12.2

20

As can be seen from the Table above the Termamyl[®]-linker-CBDEGV gives a improved liquefaction per µg enzyme/g DS compared to Termamyl[®].

25

Example 10

CBD_{CenA}-Termamyl[®] starch processing

It was investigate whether or not CBD_{CenA}-Termamyl₇₅ (Cellulomonas fimi endoglucanase A CBD and Termamyl₇₅ via a linker as described in Example 8 above) gives an improved liquefaction of starch per activity unit/g dry substance
5 compared to Termamyl₇₅ at pH 6.0 and 40 ppm Ca²⁺.

A shaking oil bath was heated to 105°C. Two starch slurries (30% DS with 40 ppm Ca⁺⁺) were prepared, the pH was adjusted to 6.0 with NaOH. The enzyme was well mixed to the slurries according to the scheme:

10

Slurry 1: CBD_{CenA}-Termamyl₇₅ 75NU/g DS starch

Slurry 2: Termamyl₇₅ 75NU/g DS starch

From each slurry four portions of 10 g each were taken.
15 Each portion were placed in an Erlenmeyer flask with screw cap. The flasks were placed in the oil bath for 8 minutes at 105°C and then 90 minutes at 95°C.

After 7 minutes and 45 seconds in the oil bath, the thermostat of the oil bath was adjusted to 95.4°C and 2 litre
20 oil at room temperature were added to the oil bath. A clock was started and samples (1 flask of each slurry) were taken after 20, 40, 60, and 90 minutes. 2 drops of 1N HCl were added to each flask to inactivate the amylase.

25 DE-determinations as function of time:

Minutes	CBD _{CenA} - Termamyl ₇₅ 75NU/g DS	Termamyl ₇₅ 75NU/g DS
20	6.1	3.9
40	8.6	6.0
60	12.0	7.7
90	15.4	10.3

As can be seen from the Table above the CBD_{CenA}-Termamyl₇₅ gives a better liquefaction per activity unit/g DS compared to Termamyl₇₅.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Novo Nordisk A/S
 (B) STREET: Novo Allé
 (C) CITY: Bagsvaerd
 (E) COUNTRY: Denmark
 (F) POSTAL CODE (ZIP): DK-2880
 (G) TELEPHONE: +45 4444 8888
 (H) TELEFAX: +45 4449 3256

(ii) TITLE OF INVENTION: Hybrid enzymes/Starch processing

(iii) NUMBER OF SEQUENCES: 37

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
 (B) COMPUTER: IBM PC compatible
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1203 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Bacillus agaradherens
 (B) STRAIN: AC13

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 1..1203

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG AAA AAG ATA ACT ACT ATT TTT GTC GTA TTG CTT ATG ACA GTG GCG	48
Met Lys Lys Ile Thr Thr Ile Phe Val Val Leu Leu Met Thr Val Ala	
1 5 10 15	
TTG TTC AGT ATA GGA AAC ACG ACT GCT GCT GAT AAT GAT TCA GTT GTA	96
Leu Phe Ser Ile Gly Asn Thr Thr Ala Asp Asn Asp Ser Val Val	
20 25 30	
GAA GAA CAT GGG CAA TTA AGT ATT AGT AAC GGT GAA TTA GTC AAT GAA	144
Glu Glu His Gly Gln Leu Ser Ile Ser Asn Gly Glu Val Asn Glu	
35 40 45	
CGA GGC GAA CAA GTT CAG TTA AAA GGG ATG AGT TCC CAT GGT TTG CAA	192
Arg Gly Glu Gln Val Gln Leu Lys Gly Met Ser Ser His Gly Leu Gln	
50 55 60	
TGG TAC GGT CAA TTT GTA AAC TAT GAA AGT ATG AAA TGG CTA AGA GAT	240
Trp Tyr Gly Gln Phe Val Asn Tyr Glu Ser Met Lys Trp Leu Arg Asp	
65 70 75 80	
GAT TGG GGA ATA AAT GTA TTC CGA GCA GCA ATG TAT ACC TCT TCA GGA	288
Asp Trp Gly Ile Asn Val Phe Arg Ala Ala Met Tyr Thr Ser Ser Gly	
85 90 95	
GGA TAT ATT GAT GAT CCA TCA GTA AAG GAA AAA GTA AAA GAG GCT GTT	336
Gly Tyr Ile Asp Asp Pro Ser Val Lys Glu Lys Val Lys Glu Ala Val	
100 105 110	
GAA GCT GCG ATA GAC CTT GAT ATA TAT GTG ATC ATT GAT TGG CAT ATC	384
Glu Ala Ala Ile Asp Leu Asp Ile Tyr Val Ile Ile Asp Trp His Ile	
115 120 125	
CTT TCA GAC AAT GAC CCA AAT ATA TAT AAA GAA GAA GCG AAG GAT TTC	432
Leu Ser Asp Asn Asp Pro Asn Ile Tyr Lys Glu Glu Ala Lys Asp Phe	
130 135 140	

TTT GAT GAA ATG TCA GAG TTG TAT GGA GAC TAT CCG AAT GTG ATA TAC Phe Asp Glu Met Ser Glu Leu Tyr Gly Asp Tyr Pro Asn Val Ile Tyr 145 150 155 160	480
GAA ATT GCA AAT GAA CCG AAT GGT AGT GAT GTT ACG TGG GGC AAT CAA Glu Ile Ala Asn Glu Pro Asn Gly Ser Asp Val Thr Trp Gly Asn Gln 165 170 175	528
ATA AAA CCG TAT GCA GAG GAA GTC ATT CCG ATT ATT CGT AAC AAT GAC Ile Lys Pro Tyr Ala Glu Glu Val Ile Pro Ile Ile Arg Asn Asn Asp 180 185 190	576
CCT AAT AAC ATT ATT ATT GTA GGT ACA GGT ACA TGG AGT CAG GAT GTC Pro Asn Asn Ile Ile Ile Val Gly Thr Gly Thr Trp Ser Gln Asp Val 195 200 205	624
CAT CAT GCA GCT GAT AAT CAG CTT GCA GAT CCT AAC GTC ATG TAT GCA His His Ala Ala Asp Asn Gln Leu Ala Asp Pro Asn Val Met Tyr Ala 210 215 220	672
TTT CAT TTT TAT GCA GGG ACA CAT GGT CAA AAT TTA CGA GAC CAA GTA Phe His Phe Tyr Ala Gly Thr His Gly Gln Asn Leu Arg Asp Gln Val 225 230 235 240	720
GAT TAT GCA TTA GAT CAA GGA GCA GCG ATA TTT GTT AGT GAA TGG GGA Asp Tyr Ala Leu Asp Gln Gly Ala Ala Ile Phe Val Ser Glu Trp Gly 245 250 255	768
ACA AGT GCA GCT ACA GGT GAT GGT GGC GTG TTT TTA GAT GAA GCA CAA Thr Ser Ala Ala Thr Gly Asp Gly Gly Val Phe Leu Asp Glu Ala Gln 260 265 270	816
GTG TGG ATT GAC TTT ATG GAT GAA AGA AAT TTA AGC TGG GCC AAC TGG Val Trp Ile Asp Phe Met Asp Glu Arg Asn Leu Ser Trp Ala Asn Trp 275 280 285	864
TCT CTA ACG CAT AAA GAT GAG TCA TCT GCA GCG TTA ATG CCA GGT GCA Ser Leu Thr His Lys Asp Glu Ser Ser Ala Ala Leu Met Pro Gly Ala 290 295 300	912
AAT CCA ACT GGT GGT TGG ACA GAG GCT GAA CTA TCT CCA TCT GGT ACA Asn Pro Thr Gly Gly Trp Thr Glu Ala Glu Leu Ser Pro Ser Gly Thr 305 310 315 320	960
TTT GTG AGG GAA AAA ATA AGA GAA TCA GCA TCT ATT CCG CCA AGC GAT Phe Val Arg Glu Lys Ile Arg Glu Ser Ala Ser Ile Pro Pro Ser Asp 325 330 335	1008
CCA ACA CCG CCA TCT GAT CCA GGA GAA CCG GAT CCA ACG CCC CCA AGT Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Asp Pro Thr Pro Pro Ser 340 345 350	1056
GAT CCA GGA GAG TAT CCA GCA TGG GAT CCA AAT CAA ATT TAC ACA AAT Asp Pro Gly Glu Tyr Pro Ala Trp Asp Pro Asn Gln Ile Tyr Thr Asn 355 360 365	1104
GAA ATT GTG TAC CAT AAC GGC CAG CTA TGG CAA GCA AAA TGG TGG ACA Glu Ile Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr 370 375 380	1152
CAA AAT CAA GAG CCA GGT GAC CCG TAC GGT CCG TGG GAA CCA CTC AAT Gln Asn Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Asn 385 390 395 400	1200
TAA	1203

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 400 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Lys Lys Ile Thr Thr Ile Phe Val Val Leu Leu Met Thr Val Ala
 1           5           10           15
Leu Phe Ser Ile Gly Asn Thr Thr Ala Ala Asp Asn Asp Ser Val Val
          20           25           30
Glu Glu His Gly Gln Leu Ser Ile Ser Asn Gly Glu Leu Val Asn Glu
          35           40           45
Arg Gly Glu Gln Val Gln Leu Lys Gly Met Ser Ser His Gly Leu Gln
          50           55           60
Trp Tyr Gly Gln Phe Val Asn Tyr Glu Ser Met Lys Trp Leu Arg Asp
          65           70           75           80
Asp Trp Gly Ile Asn Val Phe Arg Ala Ala Met Tyr Thr Ser Ser Gly
          85           90           95
Gly Tyr Ile Asp Asp Pro Ser Val Lys Glu Lys Val Lys Glu Ala Val
          100          105          110
Glu Ala Ala Ile Asp Leu Asp Ile Tyr Val Ile Ile Asp Trp His Ile
          115          120          125
Leu Ser Asp Asn Asp Pro Asn Ile Tyr Lys Glu Glu Ala Lys Asp Phe
          130          135          140
Phe Asp Glu Met Ser Glu Leu Tyr Gly Asp Tyr Pro Asn Val Ile Tyr
          145          150          155          160
Glu Ile Ala Asn Glu Pro Asn Gly Ser Asp Val Thr Trp Gly Asn Gln
          165          170          175
Ile Lys Pro Tyr Ala Glu Glu Val Ile Pro Ile Ile Arg Asn Asn Asp
          180          185          190
Pro Asn Asn Ile Ile Ile Val Gly Thr Gly Thr Trp Ser Gln Asp Val
          195          200          205
His His Ala Ala Asp Asn Gln Leu Ala Asp Pro Asn Val Met Tyr Ala
          210          215          220
Phe His Phe Tyr Ala Gly Thr His Gly Gln Asn Leu Arg Asp Gln Val
          225          230          235          240
Asp Tyr Ala Leu Asp Gln Gly Ala Ala Ile Phe Val Ser Glu Trp Gly
          245          250          255
Thr Ser Ala Ala Thr Gly Asp Gly Gly Val Phe Leu Asp Glu Ala Gln
          260          265          270
Val Trp Ile Asp Phe Met Asp Glu Arg Asn Leu Ser Trp Ala Asn Trp
          275          280          285
Ser Leu Thr His Lys Asp Glu Ser Ser Ala Ala Leu Met Pro Gly Ala
          290          295          300
Asn Pro Thr Gly Gly Trp Thr Glu Ala Glu Leu Ser Pro Ser Gly Thr
          305          310          315          320

```

Phe	Val	Arg	Glu	Lys 325	Ile	Arg	Glu	Ser	Ala 330	Ser	Ile	Pro	Pro	Ser 335	Asp
Pro	Thr	Pro	Pro 340	Ser	Asp	Pro	Gly	Glu 345	Pro	Asp	Pro	Thr	Pro 350	Pro	Ser
Asp	Pro	Gly 355	Glu	Tyr	Pro	Ala	Trp 360	Asp	Pro	Asn	Gln	Ile 365	Tyr	Thr	Asn
Glu	Ile 370	Val	Tyr	His	Asn	Gly 375	Gln	Leu	Trp	Gln	Ala 380	Lys	Trp	Trp	Thr
Gln 385	Asn	Gln	Glu	Pro	Gly 390	Asp	Pro	Tyr	Gly	Pro 395	Trp	Glu	Pro	Leu	Asn 400

- ```
(2) INFORMATION FOR SEQ ID NO:3:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 49 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 1 (#9555)"
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature
 (B) LOCATION: 33,36,39,42,45,48
 (D): OTHER INFORMATION: /Note N= A,G,C or T
 R= G or A
 Y= C or T
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
```

TCACAGATCC TCGCGAATTG GTGCGGCCGC GTNGTNGARG ARCAYGGNC 49

- (2) INFORMATION FOR SEQ ID NO:4:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 7 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (ii) MOLECULE TYPE: protein  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Val Val Glu Glu His Gly Gln  
5

- ```
(2) INFORMATION FOR SEQ ID NO:5:
  (i) SEQUENCE CHARACTERISTICS:
      (A) LENGTH: 19 base pairs
      (B) TYPE: nucleic acid
      (C) STRANDEDNESS: single
      (D) TOPOLOGY: linear
  (ii) MOLECULE TYPE: other nucleic acid
  (ix) FEATURE:
      (A) NAME/KEY: misc-feature:
      (B) OTHER INFORMATION: /desc = "Primer 2"
  (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
```

CAGAGCAAGAG ATTACGCGC 19

- (2) INFORMATION FOR SEQ ID NO:6:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 17 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:

(A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Reverse Primer"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTTTTCCCAG TCACGAC 17

(2) INFORMATION FOR SEQ ID NO:7:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Forward Primer"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCGGATAACA ATTCACACA GG 22

(2) INFORMATION FOR SEQ ID NO:8:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 3, #19719"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TGACCCGTAC GGTCCGTGGG 20

(2) INFORMATION FOR SEQ ID NO:9:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 4, #19720"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGCTCTTGAT TTTGTGTCCA CC 22

(2) INFORMATION FOR SEQ ID NO:10:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 71 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 5. #20887"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GTAGGCTCAG TCATATGTTA CACATTGAAA GGGGAGGAGA ATCATGAAAA AGATAACTAC 60
 TATTTTGTGTC G 71

(2) INFORMATION FOR SEQ ID NO:11:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 51 base pairs
 (B) TYPE: nucleic acid

51

- (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 6"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTACCTCGCG GGTACCAAGC GGCCGCTTAA TTGAGTGGTT CCCACGGACC G

51

- (2) INFORMATION FOR SEQ ID NO: 12:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1386 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: DNA (genomic)
 (vi) ORIGINAL SOURCE:
 (A) ORGANISM: Bacillus agaradherens
 (B) STRAIN: AC13
 (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION:1..1386
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ATG AAA AAG ATA ACT ACT ATT TTT GTC GTA TTG CTT ATG ACA GTG GCG	48
Met Lys Lys Ile Thr Thr Ile Phe Val Val Leu Leu Met Thr Val Ala	
1 5 10 15	
TTG TTC AGT ATA GGA AAC ACG ACT GCT GCT GAT AAT GAT TCA GTT GTA	96
Leu Phe Ser Ile Gly Asn Thr Thr Ala Ala Asp Asn Asp Ser Val Val	
20 25 30	
GAA GAA CAT GGG CAA TTA AGT ATT AGT AAC GGT GAA TTA GTC AAT GAA	144
Glu Glu His Gly Gln Leu Ser Ile Ser Asn Gly Glu Leu Val Asn Glu	
35 40 45	
CGA GGC GAA CAA GTT CAG TTA AAA GGG ATG AGT TCC CAT GGT TTG CAA	192
Arg Gly Glu Gln Val Gln Leu Lys Gly Met Ser Ser His Gly Leu Gln	
50 55 60	
TGG TAC GGT CAA TTT GTA AAC TAT GAA AGT ATG AAA TGG CTA AGA GAT	240
Trp Tyr Gly Gln Phe Val Asn Tyr Glu Ser Met Lys Trp Leu Arg Asp	
65 70 75 80	
GAT TGG GGA ATA AAT GTA TTC CGA GCA GCA ATG TAT ACC TCT TCA GGA	288
Asp Trp Gly Ile Asn Val Phe Arg Ala Ala Met Tyr Thr Ser Ser Gly	
85 90 95	
GGA TAT ATT GAT GAT CCA TCA GTA AAG GAA AAA GTA AAA GAG GCT GTT	336
Gly Tyr Ile Asp Asp Pro Ser Val Lys Glu Lys Val Lys Glu Ala Val	
100 105 110	
GAA GCT GCG ATA GAC CTT GAT ATA TAT GTG ATC ATT GAT TGG CAT ATC	384
Glu Ala Ala Ile Asp Leu Asp Ile Tyr Val Ile Ile Asp Trp His Ile	
115 120 125	
CTT TCA GAC AAT GAC CCA AAT ATA TAT AAA GAA GAA GCG AAG GAT TTC	432
Leu Ser Asp Asn Asp Pro Asn Ile Tyr Lys Glu Glu Ala Lys Asp Phe	
130 135 140	
TTT GAT GAA ATG TCA GAG TTG TAT GGA GAC TAT CCG AAT GTG ATA TAC	480
Phe Asp Glu Met Ser Glu Leu Tyr Gly Asp Tyr Pro Asn Val Ile Tyr	
145 150 155 160	
GAA ATT GCA AAT GAA CCG AAT GGT AGT GAT GTT ACG TGG GGC AAT CAA	528
Glu Ile Ala Asn Glu Pro Asn Gly Ser Asp Val Thr Trp Gly Asn Gln	
165 170 175	

ATA AAA CCG TAT GCA GAG GAA GTC ATT CCG ATT ATT CGT AAC AAT GAC	576
Ile Lys Pro Tyr Ala Glu Glu Val Ile Pro Ile Ile Arg Asn Asn Asp	
180 185 190	
CCT AAT AAC ATT ATT ATT GTA GGT ACA GGT ACA TGG AGT CAG GAT GTC	624
Pro Asn Asn Ile Ile Ile Val Gly Thr Gly Thr Trp Ser Gln Asp Val	
195 200 205	
CAT CAT GCA GCT GAT AAT CAG CTT GCA GAT CCT AAC GTC ATG TAT GCA	672
His His Ala Ala Asp Asn Gln Leu Ala Asp Pro Asn Val Met Tyr Ala	
210 215 220	
TTT CAT TTT TAT GCA GGG ACA CAT GGT CAA AAT TTA CGA GAC CAA GTA	720
Phe His Phe Tyr Ala Gly Thr His Gly Gln Asn Leu Arg Asp Gln Val	
225 230 235 240	
GAT TAT GCA TTA GAT CAA GCA GCA GCG ATA TTT GTT AGT GAA TGG GGA	768
Asp Tyr Ala Leu Asp Gln Gly Ala Ala Ile Phe Val Ser Glu Trp Gly	
245 250 255	
ACA AGT GCA GCT ACA GGT GAT GGT GGC GTG TTT TTA GAT GAA GCA CAA	816
Thr Ser Ala Ala Thr Gly Asp Gly Gly Val Phe Leu Asp Glu Ala Gln	
260 265 270	
GTG TGG ATT GAC TTT ATG GAT GAA AGA AAT TTA AGC TGG GCC AAC TGG	864
Val Trp Ile Asp Phe Met Asp Glu Arg Asn Leu Ser Trp Ala Asn Trp	
275 280 285	
TCT CTA ACG CAT AAA GAT GAG TCA TCT GCA GCG TTA ATG CCA GGT GCA	912
Ser Leu Thr His Lys Asp Glu Ser Ser Ala Ala Leu Met Pro Gly Ala	
290 295 300	
AAT CCA ACT GGT GGT TGG ACA GAG GCT GAA CTA TCT CCA TCT GGT ACA	960
Asn Pro Thr Gly Gly Trp Thr Glu Ala Glu Leu Ser Pro Ser Gly Thr	
305 310 315 320	
TTT GTG AGG GAA AAA ATA AGA GAA TCA GCA TCT ATT CCG CCA AGC GAT	1008
Phe Val Arg Glu Lys Ile Arg Glu Ser Ala Ser Ile Pro Pro Ser Asp	
325 330 335	
CCA ACA CCG CCA TCT GAT CCA GGA GAA CCG GAT CCA ACG CCC CCA AGT	1056
Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Asp Pro Thr Pro Ser	
340 345 350	
GAT CCA GGA AAG TAT CCA GCA TGG GAT CCA AAT CAA ATT TAC ACA AAT	1104
Asp Pro Gly Lys Tyr Pro Ala Trp Asp Pro Asn Gln Ile Tyr Thr Asn	
355 360 365	
GAA ATT GTG TAC CAT AAC GGC CAG CTA TGG CAA GCA AAA TGG TGG ACA	1152
Glu Ile Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr	
370 375 380	
CAA AAT CAA GAG CCA GGT GAC CCG TAC GGT CCG TGG GAA CCA CTC AAA	1200
Gln Asn Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Lys	
385 390 395 400	
TCT GAT CCA GAT TCA GGA GAA CCG GAT CCA ACG CCC CCA AGT GAT CCA	1248
Ser Asp Pro Asp Ser Gly Glu Pro Asp Pro Thr Pro Pro Ser Asp Pro	
405 410 415	
GGA GAA TAT CCA GCA TGG GAC CCA ACG CAA ATT TAC ACA GAT GAA ATT	1296
Gly Glu Tyr Pro Ala Trp Asp Pro Thr Gln Ile Tyr Thr Asp Glu Ile	
420 425 430	
GTG TAC CAT AAC GGC CAG CTA TGG CAA GCC AAA TGG TGG ACA CAA AAT	1344
Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr Gln Asn	
435 440 445	

CAA GAG CCA GGT GAC CCA TAC GGT CCG TGG GAA CCA CTC AAT 1386
 Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Asn
 450 455 460

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 462 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Lys Lys Ile Thr Thr Ile Phe Val Val Leu Leu Met Thr Val Ala
 1 5 10 15
 Leu Phe Ser Ile Gly Asn Thr Thr Ala Ala Asp Asn Asp Ser Val Val
 20 25 30
 Glu Glu His Gly Gln Leu Ser Ile Ser Asn Gly Glu Leu Val Asn Glu
 35 40 45
 Arg Gly Glu Gln Val Gln Leu Lys Gly Met Ser Ser His Gly Leu Gln
 50 55 60
 Trp Tyr Gly Gln Phe Val Asn Tyr Glu Ser Met Lys Trp Leu Arg Asp
 65 70 75 80
 Asp Trp Gly Ile Asn Val Phe Arg Ala Ala Met Tyr Thr Ser Ser Gly
 85 90 95
 Gly Tyr Ile Asp Asp Pro Ser Val Lys Glu Lys Val Lys Glu Ala Val
 100 105 110
 Glu Ala Ala Ile Asp Leu Asp Ile Tyr Val Ile Ile Asp Trp His Ile
 115 120 125
 Leu Ser Asp Asn Asp Pro Asn Ile Tyr Lys Glu Glu Ala Lys Asp Phe
 130 135 140
 Phe Asp Glu Met Ser Glu Leu Tyr Gly Asp Tyr Pro Asn Val Ile Tyr
 145 150 155 160
 Glu Ile Ala Asn Glu Pro Asn Gly Ser Asp Val Thr Trp Gly Asn Gln
 165 170 175
 Ile Lys Pro Tyr Ala Glu Glu Val Ile Pro Ile Ile Arg Asn Asn Asp
 180 185 190
 Pro Asn Asn Ile Ile Ile Val Gly Thr Gly Thr Trp Ser Gln Asp Val
 195 200 205
 His His Ala Ala Asp Asn Gln Leu Ala Asp Pro Asn Val Met Tyr Ala
 210 215 220
 Phe His Phe Tyr Ala Gly Thr His Gly Gln Asn Leu Arg Asp Gln Val
 225 230 235 240
 Asp Tyr Ala Leu Asp Gln Gly Ala Ala Ile Phe Val Ser Glu Trp Gly
 245 250 255
 Thr Ser Ala Ala Thr Gly Asp Gly Gly Val Phe Leu Asp Glu Ala Gln
 260 265 270
 Val Trp Ile Asp Phe Met Asp Glu Arg Asn Leu Ser Trp Ala Asn Trp
 275 280 285
 Ser Leu Thr His Lys Asp Glu Ser Ser Ala Ala Leu Met Pro Gly Ala

54

290	295	300
Asn Pro Thr Gly Gly Trp Thr Glu Ala Glu Leu Ser Pro Ser Gly Thr		
305	310	315 320
Phe Val Arg Glu Lys Ile Arg Glu Ser Ala Ser Ile Pro Pro Ser Asp		
	325	330 335
Pro Thr Pro Pro Ser Asp Pro Gly Glu Pro Asp Pro Thr Pro Pro Ser		
	340	345 350
Asp Pro Gly Lys Tyr Pro Ala Trp Asp Pro Asn Gln Ile Tyr Thr Asn		
	355	360 365
Glu Ile Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr		
	370	375 380
Gln Asn Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Lys		
	385	390 395 400
Ser Asp Pro Asp Ser Gly Glu Pro Asp Pro Thr Pro Pro Ser Asp Pro		
	405	410 415
Gly Glu Tyr Pro Ala Trp Asp Pro Thr Gln Ile Tyr Thr Asp Glu Ile		
	420	425 430
Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr Gln Asn		
	435	440 445
Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Asn		
	450	455 460

- (2) INFORMATION FOR SEQ ID NO: 14:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 7, #100084"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CCTCGCGAGG TACCAGCGGC CGCGTACCAC CAATTAAGTA TGGTAC

46

- (2) INFORMATION FOR SEQ ID NO: 15:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 35 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 8, #5289"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

GCTTTACGCC CGATTGCTGA CGCTG

35

- (2) INFORMATION FOR SEQ ID NO: 16:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 51 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
- (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 9, #26748"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

GCGATGAGAC GCGCGGCCGC CTATCTTTGA ACATAAATTG AAACGGATCC G 51

- (2) INFORMATION FOR SEQ ID NO: 17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 52 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 10, #110150A"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

GCTGCAGGAT CCGTTTCAAT TTATGTTCAA AGATCTGATC CAGATTCAGG AG 52

- (2) INFORMATION FOR SEQ ID NO: 18:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 46 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 11, #100084"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CCTCGCGAGG TACCAGCGGC CGCGTACCAC CAATTAAGTA TGGTAC 46

- (2) INFORMATION FOR SEQ ID NO: 19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1725 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Hybrid"
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..1725
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

ATG AAA CAA CAA AAA CGG CTT TAC GCC CGA TTG CTG ACG CTG TTA TTT 48
Met Lys Gln Gln Lys Arg Leu Tyr Ala Arg Leu Leu Thr Leu Leu Phe
1 5 10 15

GCG CTC ATC TTC TTG CTG CCT CAT TCT GCA GCA GCG GCG GCA AAT CTT 96
Ala Leu Ile Phe Leu Leu Pro His Ser Ala Ala Ala Ala Ala Asn Leu
20 25 30

AAT GGG ACG CTG ATG CAG TAT TTT GAA TGG TAC ATG CCC AAT GAC GGC 144
Asn Gly Thr Leu Met Gln Tyr Phe Glu Trp Tyr Met Pro Asn Asp Gly
35 40 45

CAA CAT TGG AAG CGT TTG CAA AAC GAC TCG GCA TAT TTG GCT GAA CAC 192
Gln His Trp Lys Arg Leu Gln Asn Asp Ser Ala Tyr Leu Ala Glu His
50 55 60

GGT ATT ACT GCC GTC TGG ATT CCC CCG GCA TAT AAG GGA ACG AGC CAA 240
Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Tyr Lys Gly Thr Ser Gln

56

65	70					75					80					
GCG GAT GTG GGC TAC GGT GCT TAC GAC CTT TAT GAT TTA GGG GAG TTT	Ala Asp Val Gly Tyr 85					Gly Ala Tyr Asp Leu 90					Gly Glu Phe 95					288
CAT CAA AAA GGG ACG GTT CGG ACA AAG TAC GGC ACA AAA GGA GAG CTG	His Gln Lys 100					Thr Val Arg Thr Lys 105					Gly Thr Lys Glu 110					336
CAA TCT GCG ATC AAA AGT CTT CAT TCC CGC GAC ATT AAC GTT TAC GGG	Gln Ser 115					Leu His Ser Arg Asp 120					Ile Asn Val Tyr 125					384
GAT GTG GTC ATC AAC CAC AAA GGC GGC GCT GAT GCG ACC GAA GAT GTA	Asp Val 130					Lys Gly Thr Lys 135					Ala Thr Glu Asp 140					432
ACC GCG GTT GAA GTC GAT CCC GCT GAC CGC AAC CGC GTA ATC TCA GGA	Thr Ala Val Glu Val 145					Pro Ala Asp Arg Asn 155					Arg Val Ile Ser 160					480
GAA CAC CTA ATT AAA GCC TGG ACA CAT TTT CAT TTT CCG GGG GCC GGC	Glu His Leu Ile Lys 165					Ala Trp Thr His Phe 170					Pro Gly Ala Gly 175					528
AGC ACA TAC AGC GAT TTT AAA TGG CAT TGG TAC CAT TTT GAC GGA ACC	Ser Thr Tyr Ser Asp 180					Phe Lys Trp His Trp 185					His Phe Asp Gly 190					576
GAT TGG GAC GAG TCC CGA AAG CTG AAC CGC ATC TAT AAG TTT CAA GGA	Asp Trp Asp Glu Ser Arg Lys 195					Leu Asn Arg Ile Tyr 205					Lys Phe Gln Gly 210					624
AAG GCT TGG GAT TGG GAA GTT TCC AAT GAA AAC GGC AAC TAT GAT TAT	Lys Ala Trp Asp Trp Glu Val Ser Asn Glu Asn Gly Asn Tyr Asp Tyr 210					215					220					672
TTG ATG TAT GCC GAC ATC GAT TAT GAC CAT CCT GAT GTC GCA GCA GAA	Leu Met Tyr Ala Asp Ile Asp Tyr Asp His Pro Asp Val Ala Ala Glu 225					230					235					720
ATT AAG AGA TGG GGC ACT TGG TAT GCC AAT GAA CTG CAA TTG GAC GGA	Ile Lys Arg Trp Gly Thr Trp Tyr Ala Asn Glu Leu Gln Leu Asp Gly 245					250					255					768
AAC CGT CTT GAT GCT GTC AAA CAC ATT AAA TTT TCT TTT TTG CGG GAT	Asn Arg Leu Asp Ala Val Lys His Ile Lys Phe Ser Phe Leu Arg Asp 260					265					270					816
TGG GTT AAT CAT GTC AGG GAA AAA ACG GGC AAG GAA ATG TTT ACG GTA	Trp Val Asn His Val Arg Glu Lys Thr Gly Lys Glu Met Phe Thr Val 275					280					285					864
GCT GAA TAT TGG CAG AAT GAC TTG GGC GCG CTG GAA AAC TAT TTG AAC	Ala Glu Tyr Trp Gln Asn Asp Leu Gly Ala Leu Glu Asn Tyr Leu Asn 290					295					300					912
AAA ACA AAT TTT AAT CAT TCA GTG TTT GAC GTG CCG CTT CAT TAT CAG	Lys Thr Asn Phe Asn His Ser Val Phe Asp Val Pro Leu His Tyr Gln 305					310					315					960
TTC CAT GCT GCA TCG ACA CAG GGA GGC GGC TAT GAT ATG AGG AAA TTG	Phe His Ala Ala Ser Thr Gln Gly Gly Gly Tyr Asp Met Arg Lys Leu 325					330					335					1008
CTG AAC GGT ACG GTC GTT TCC AAG CAT CCG TTG AAA TCG GTT ACA TTT	Leu Asn Gly Thr Val Val Ser Lys His Pro Leu Lys Ser Val Thr Phe 1056															

340					345					350						
GTC Val	GAT Asp	AAC Asn 355	CAT His	GAT Asp	ACA Thr	CAG Gln	CCG Pro 360	GGG Gly	CAA Gln	TCG Ser	CTT Leu	GAG Glu 365	TCG Ser	ACT Thr	GTC Val	1104
CAA Gln	ACA Thr 370	TGG Trp	TTT Phe	AAG Lys	CCG Pro	CTT Leu 375	GCT Ala	TAC Tyr	GCT Ala	TTT Phe	ATT Ile 380	CTC Leu	ACA Thr	AGG Arg	GAA Glu	1152
TCT Ser 385	GGA Gly	TAC Tyr	CCT Pro	CAG Gln	GTT Val 390	TTC Phe	TAC Tyr	GGG Gly	GAT Asp	ATG Met 395	TAC Tyr	GGG Gly	ACG Thr	AAA Lys	GGA Gly 400	1200
GAC Asp	TCC Ser	CAG Gln	CGC Arg	GAA Glu 405	ATT Ile	CCT Pro	GCC Ala	TTG Leu	AAA Lys 410	CAC His	AAA Lys	ATT Ile	GAA Glu	CCG Pro 415	ATC Ile	1248
TTA Leu	AAA Lys	GCG Ala	AGA Arg 420	AAA Lys	CAG Gln	TAT Tyr	GCG Ala	TAC Tyr 425	GGA Gly	GCA Ala	CAG Gln	CAT His	GAT Asp 430	TAT Tyr	TTC Phe	1296
GAC Asp	CAC His	CAT His 435	GAC Asp	ATT Ile	GTC Val	GGC Gly	TGG Trp 440	ACA Thr	AGG Arg	GAA Glu	GGC Gly	GAC Asp 445	AGC Ser	TCG Ser	GTT Val	1344
GCA Ala	AAT Asn 450	TCA Ser	GGT Gly	TTG Leu	GCG Ala	GCA Ala 455	TTA Leu	ATA Ile	ACA Thr	GAC Asp	GGA Gly 460	CCC Pro	GGT Gly	GGG Gly	GCA Ala	1392
AAG Lys 465	CGA Arg	ATG Met	TAT Tyr	GTC Val	GGC Gly 470	CGG Arg	CAA Gln	AAC Asn	GCC Ala	GGT Gly 475	GAG Glu	ACA Thr	TGG Trp	CAT His	GAC Asp 480	1440
ATT Ile	ACC Thr	GGA Gly	AAC Asn	CGT Arg 485	TCG Ser	GAG Glu	CCG Pro	GTT Val	GTC Val 490	ATC Ile	AAT Asn	TCG Ser	GAA Glu	GGC Gly 495	TGG Trp	1488
GGA Gly	GAG Glu	TTT Phe	CAC His 500	GTA Val	AAC Asn	GGC Gly	GGA Gly	TCC Ser 505	GTT Val	TCA Ser	ATT Ile	TAT Tyr	GTT Val 510	CAA Gln	AGA Arg	1536
TCT Ser	GAT Asp	CCA Pro 515	GAT Asp	TCA Ser	GGA Gly	GAA Glu	CCG Pro 520	GAT Asp	CCA Pro	ACG Thr	CCC Pro	CCA Pro 525	AGT Ser	GAT Asp	CCA Pro	1584
GGA Gly	GAA Glu 530	TAT Tyr	CCA Pro	GCA Ala	TGG Trp	GAC Asp 535	CCA Pro	ACG Thr	CAA Gln	ATT Ile	TAC Tyr 540	ACA Thr	GAT Asp	GAA Glu	ATT Ile	1632
GTG Val 545	TAC Tyr	CAT His	AAC Asn	GGC Gly	CAG Gln 550	CTA Leu	TGG Trp	CAA Gln	GCC Ala	AAA Lys 555	TGG Trp	TGG Trp	ACA Thr	CAA Gln	AAT Asn 560	1680
CAA Gln	GAG Glu	CCA Pro	GGT Gly	GAC Asp 565	CCA Pro	TAC Tyr	GGT Gly	CCG Pro	TGG Trp 570	GAA Glu	CCA Pro	CTC Leu	AAT Asn	TAA * 575		1725

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 575 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Lys Gln Gln Lys Arg Leu Tyr Ala Arg Leu Leu Thr Leu Leu Phe
 1 5 10 15

Ala Leu Ile Phe Leu Leu Pro His Ser Ala Ala Ala Ala Ala Asn Leu
 20 25 30
 Asn Gly Thr Leu Met Gln Tyr Phe Glu Trp Tyr Met Pro Asn Asp Gly
 35 40 45
 Gln His Trp Lys Arg Leu Gln Asn Asp Ser Ala Tyr Leu Ala Glu His
 50 55 60
 Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Tyr Lys Gly Thr Ser Gln
 65 70 75 80
 Ala Asp Val Gly Tyr Gly Ala Tyr Asp Leu Tyr Asp Leu Gly Glu Phe
 85 90 95
 His Gln Lys Gly Thr Val Arg Thr Lys Tyr Gly Thr Lys Gly Glu Leu
 100 105 110
 Gln Ser Ala Ile Lys Ser Leu His Ser Arg Asp Ile Asn Val Tyr Gly
 115 120 125
 Asp Val Val Ile Asn His Lys Gly Gly Ala Asp Ala Thr Glu Asp Val
 130 135 140
 Thr Ala Val Glu Val Asp Pro Ala Asp Arg Asn Arg Val Ile Ser Gly
 145 150 155 160
 Glu His Leu Ile Lys Ala Trp Thr His Phe His Phe Pro Gly Ala Gly
 165 170 175
 Ser Thr Tyr Ser Asp Phe Lys Trp His Trp Tyr His Phe Asp Gly Thr
 180 185 190
 Asp Trp Asp Glu Ser Arg Lys Leu Asn Arg Ile Tyr Lys Phe Gln Gly
 195 200 205
 Lys Ala Trp Asp Trp Glu Val Ser Asn Glu Asn Gly Asn Tyr Asp Tyr
 210 215 220
 Leu Met Tyr Ala Asp Ile Asp Tyr Asp His Pro Asp Val Ala Ala Glu
 225 230 235 240
 Ile Lys Arg Trp Gly Thr Trp Tyr Ala Asn Glu Leu Gln Leu Asp Gly
 245 250 255
 Asn Arg Leu Asp Ala Val Lys His Ile Lys Phe Ser Phe Leu Arg Asp
 260 265 270
 Trp Val Asn His Val Arg Glu Lys Thr Gly Lys Glu Met Phe Thr Val
 275 280 285
 Ala Glu Tyr Trp Gln Asn Asp Leu Gly Ala Leu Glu Asn Tyr Leu Asn
 290 295 300
 Lys Thr Asn Phe Asn His Ser Val Phe Asp Val Pro Leu His Tyr Gln
 305 310 315 320
 Phe His Ala Ala Ser Thr Gln Gly Gly Gly Tyr Asp Met Arg Lys Leu
 325 330 335
 Leu Asn Gly Thr Val Val Ser Lys His Pro Leu Lys Ser Val Thr Phe
 340 345 350
 Val Asp Asn His Asp Thr Gln Pro Gly Gln Ser Leu Glu Ser Thr Val
 355 360 365
 Gln Thr Trp Phe Lys Pro Leu Ala Tyr Ala Phe Ile Leu Thr Arg Glu

59

```

      370              375              380
Ser Gly Tyr Pro Gln Val Phe Tyr Gly Asp Met Tyr Gly Thr Lys Gly
385              390              395              400
Asp Ser Gln Arg Glu Ile Pro Ala Leu Lys His Lys Ile Glu Pro Ile
              405              410              415
Leu Lys Ala Arg Lys Gln Tyr Ala Tyr Gly Ala Gln His Asp Tyr Phe
              420              425              430
Asp His His Asp Ile Val Gly Trp Thr Arg Glu Gly Asp Ser Ser Val
              435              440              445
Ala Asn Ser Gly Leu Ala Ala Leu Ile Thr Asp Gly Pro Gly Gly Ala
              450              455              460
Lys Arg Met Tyr Val Gly Arg Gln Asn Ala Gly Glu Thr Trp His Asp
465              470              475              480
Ile Thr Gly Asn Arg Ser Glu Pro Val Val Ile Asn Ser Glu Gly Trp
              485              490              495
Gly Glu Phe His Val Asn Gly Gly Ser Val Ser Ile Tyr Val Gln Arg
              500              505              510
Ser Asp Pro Asp Ser Gly Glu Pro Asp Pro Thr Pro Pro Ser Asp Pro
              515              520              525
Gly Glu Tyr Pro Ala Trp Asp Pro Thr Gln Ile Tyr Thr Asp Glu Ile
              530              535              540
Val Tyr His Asn Gly Gln Leu Trp Gln Ala Lys Trp Trp Thr Gln Asn
545              550              555              560
Gln Glu Pro Gly Asp Pro Tyr Gly Pro Trp Glu Pro Leu Asn *
              565              570              575

```

2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(ix) FEATURE:

(a) NAME/KEY: misc-feature

(d) OTHER INFORMATION: /desc = "Linker"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```

Ser Asp Pro Asp Ser Gly Glu Pro Asp Pro Thr Pro Pro Ser Asp Pro Gly
      5              10              15

```

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

(A) NAME/KEY: misc-feature:

(B) OTHER INFORMATION: /desc = "Primer 12, #114135"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GCTGCAGGAT CCGTTTCAAT TTATGTTCAA AGATCTCCAA CTCCTGCCCC ATCTCAAAGC 60

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 50 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 13, #110151"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GCGATGAGAC GCGCGGCCGC TACTACCACT CAACATTAAAC AGGACCTGAG

50

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2346 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
 - (A) DESCRIPTION: /desc = "Hybrid"
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2346
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

ATG AAA CAA CAA AAA CGG CTT TAC GCC CGA TTG CTG ACG CTG TTA TTT	48
Met Lys Gln Gln Lys Arg Leu Tyr Ala Arg Leu Leu Thr Leu Leu Phe	
1 5 10 15	
GCG CTC ATC TTC TTG CTG CCT CAT TCT GCA GCA GCG GCG GCA AAT CTT	96
Ala Leu Ile Phe Leu Leu Pro His Ser Ala Ala Ala Ala Asn Leu	
20 25 30	
AAT GGG ACG CTG ATG CAG TAT TTT GAA TGG TAC ATG CCC AAT GAC GGC	144
Asn Gly Thr Leu Met Gln Tyr Phe Glu Trp Tyr Met Pro Asn Asp Gly	
35 40 45	
CAA CAT TGG AAG CGT TTG CAA AAC GAC TCG GCA TAT TTG GCT GAA CAC	192
Gln His Trp Lys Arg Leu Gln Asn Asp Ser Ala Tyr Leu Ala Glu His	
50 55 60	
GGT ATT ACT GCC GTC TGG ATT CCC CCG GCA TAT AAG GGA ACG AGC CAA	240
Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Tyr Lys Gly Thr Ser Gln	
65 70 75 80	
GCG GAT GTG GGC TAC GGT GCT TAC GAC CTT TAT GAT TTA GGG GAG TTT	288
Ala Asp Val Gly Tyr Gly Ala Tyr Asp Leu Tyr Asp Leu Gly Glu Phe	
85 90 95	
CAT CAA AAA GGG ACG GTT CGG ACA AAG TAC GGC ACA AAA GGA GAG CTG	336
His Gln Lys Gly Thr Val Arg Thr Lys Tyr Gly Thr Lys Gly Glu Leu	
100 105 110	
CAA TCT GCG ATC AAA AGT CTT CAT TCC CGC GAC ATT AAC GTT TAC GGG	384
Gln Ser Ala Ile Lys Ser Leu His Ser Arg Asp Ile Asn Val Tyr Gly	
115 120 125	
GAT GTG GTC ATC AAC CAC AAA GGC GGC GCT GAT GCG ACC GAA GAT GTA	432
Asp Val Val Ile Asn His Lys Gly Gly Ala Asp Ala Thr Glu Asp Val	
130 135 140	
ACC GCG GTT GAA GTC GAT CCC GCT GAC CGC AAC CGC GTA ATC TCA GGA	480
Thr Ala Val Glu Val Asp Pro Ala Asp Arg Asn Arg Val Ile Ser Gly	
145 150 155 160	
GAA CAC CTA ATT AAA GCC TGG ACA CAT TTT CAT TTT CCG GGG GCC GGC	528
Glu His Leu Ile Lys Ala Trp Thr His Phe His Phe Pro Gly Ala Gly	
165 170 175	

AGC Ser	ACA Thr	TAC Tyr	AGC Ser 180	GAT Asp	TTT Phe	AAA Lys	TGG Trp	CAT His 185	TGG Trp	TAC Tyr	CAT His	TTT Phe 190	GAC Asp 190	GGA Gly	ACC Thr	576
GAT Asp	TGG Trp	GAC Asp 195	GAG Glu	TCC Ser	CGA Arg	AAG Lys	CTG Leu 200	AAC Asn	CGC Arg	ATC Ile	TAT Tyr	AAG Lys 205	TTT Phe	CAA Gln	GGA Gly	624
AAG Lys	GCT Ala 210	TGG Trp	GAT Asp	TGG Trp	GAA Glu	GTT Val 215	TCC Ser	AAT Asn	GAA Glu	AAC Asn	GGC Gly 220	AAC Asn	TAT Tyr	GAT Asp	TAT Tyr	672
TTG Leu 225	ATG Met	TAT Tyr	GCC Ala	GAC Asp	ATC Ile 230	GAT Asp	TAT Tyr	GAC Asp	CAT His	CCT Pro 235	GAT Asp	GTC Val	GCA Ala	GCA Ala	GAA Glu 240	720
ATT Ile	AAG Lys	AGA Arg	TGG Trp	GGC Gly 245	ACT Thr	TGG Trp	TAT Tyr	GCC Ala 250	AAT Asn	GAA Glu	CTG Leu	CAA Gln	TTG Leu	GAC Asp 255	GGA Gly	768
AAC Asn	CGT Arg	CTT Leu 260	GAT Asp	GCT Ala	GTC Val	AAA Lys	CAC His 265	ATT Ile 265	AAA Lys	TTT Phe	TCT Ser	TTT Phe	TTG Leu 270	CGG Arg	GAT Asp	816
TGG Trp	GTT Val 275	AAT Asn	CAT His	GTC Val	AGG Arg	GAA Glu	AAA Lys 280	ACG Thr	GGG Gly	AAG Lys	GAA Glu	ATG Met 285	TTT Phe	ACG Thr	GTA Val	864
GCT Ala	GAA Glu 290	TAT Tyr	TGG Trp	CAG Gln	AAT Asn	GAC Asp 295	TTG Leu	GGC Gly	GCG Ala	CTG Leu	GAA Glu 300	AAC Asn	TAT Tyr	TTG Leu	AAC Asn	912
AAA Lys 305	ACA Thr	AAT Asn	TTT Phe	AAT Asn	CAT His 310	TCA Ser	GTG Val	TTT Phe	GAC Asp 315	GTG Val 315	CCG Pro	CTT Leu	CAT His	TAT Tyr	CAG Gln 320	960
TTC Phe	CAT His	GCT Ala	GCA Ala	TCG Ser 325	ACA Thr	CAG Gln	GGA Gly	GGC Gly 330	GGC Gly	TAT Tyr	GAT Asp	ATG Met	AGG Arg	AAA Lys 335	TTG Leu	1008
CTG Leu	AAC Asn	GGT Gly	ACG Thr 340	GTC Val	GTT Val	TCC Ser	AAG Lys	CAT His 345	CCG Pro	TTG Leu	AAA Lys	TCG Ser	GTT Val 350	ACA Thr	TTT Phe	1056
GTC Val	GAT Asp	AAC Asn 355	CAT His	GAT Asp	ACA Thr	CAG Gln	CCG Pro 360	GGG Gly	CAA Gln	TCG Ser	CTT Leu	GAG Glu 365	TCG Ser	ACT Thr	GTC Val	1104
CAA Gln 370	ACA Thr	TGG Trp	TTT Phe	AAG Lys	CCG Pro	CTT Leu 375	GCT Ala	TAC Tyr	GCT Ala	TTT Phe	ATT Ile 380	CTC Leu	ACA Thr	AGG Arg	GAA Glu	1152
TCT Ser 385	GGA Gly	TAC Tyr	CCT Pro	CAG Gln	GTT Val 390	TTC Phe	TAC Tyr	GGG Gly	GAT Asp	ATG Met 395	TAC Tyr	GGG Gly	ACG Thr	AAA Lys	GGA Gly 400	1200
GAC Asp	TCC Ser	CAG Gln	CGC Arg	GAA Glu 405	ATT Ile	CCT Pro	GCC Ala	TTG Leu	AAA Lys 410	CAC His	AAA Lys	ATT Ile	GAA Glu 415	CCG Pro	ATC Ile	1248
TTA Leu	AAA Lys	GCG Ala	AGA Arg 420	AAA Lys	CAG Gln	TAT Tyr	GCG Ala 425	TAC Tyr	GGA Gly	GCA Ala	CAG Gln	CAT His 430	GAT Asp	TAT Tyr	TTC Phe	1296
GAC Asp	CAC His	CAT His 435	GAC Asp	ATT Ile	GTC Val	GGC Gly	TGG Trp 440	ACA Thr	AGG Arg	GAA Glu	GGC Gly	GAC Asp 445	AGC Ser	TCG Ser	GTT Val	1344

GCA Ala	AAT Asn	TCA Ser	GGT Gly	TTG Leu	GCG Ala	GCA Ala	TTA Leu	ATA Ile	ACA Thr	GAC Asp	GGA Gly	CCC Pro	GGT Gly	GGG Gly	GCA Ala	1392
450						455					460					
AAG Lys	CGA Arg	ATG Met	TAT Tyr	GTC Val	GGC Gly	CGG Arg	CAA Gln	AAC Asn	GCC Ala	GGT Gly	GAG Glu	ACA Thr	TGG Trp	CAT His	GAC Asp	1440
465					470					475					480	
ATT Ile	ACC Thr	GGA Gly	AAC Asn	CGT Arg	TCG Ser	GAG Glu	CCG Pro	GTT Val	GTC Val	ATC Ile	AAT Asn	TCG Ser	GAA Glu	GGC Gly	TGG Trp	1488
				485					490					495		
GGA Gly	GAG Glu	TTT Phe	CAC His	GTA Val	AAC Asn	GGC Gly	GGA Gly	TCC Ser	GTT Val	TCA Ser	ATT Ile	TAT Tyr	GTT Val	CAA Gln	AGA Arg	1536
			500					505					510			
TCT Ser	CCA Pro	ACT Thr	CCT Pro	GCC Ala	CCA Pro	TCT Ser	CAA Gln	AGC Ser	CCA Pro	ATT Ile	AGA Arg	AGA Arg	GAT Asp	GCA Ala	TTT Phe	1584
		515					520					525				
TCA Ser	ATA Ile	ATC Ile	GAA Glu	GCG Ala	GAA Glu	GAA Glu	TAT Tyr	AAC Asn	AGC Ser	ACA Thr	AAT Asn	TCC Ser	TCC Ser	ACT Thr	TTA Leu	1632
		530				535						540				
CAA Gln	GTG Val	ATT Ile	GGA Gly	ACG Thr	CCA Pro	AAT Asn	AAT Asn	GGC Gly	AGA Arg	GGA Gly	ATT Ile	GGT Gly	TAT Tyr	ATT Ile	GAA Glu	1680
545					550					555					560	
AAT Asn	GGT Gly	AAT Asn	ACC Thr	GTA Val	ACT Thr	TAC Tyr	AGC Ser	AAT Asn	ATA Ile	GAT Asp	TTT Phe	GGT Gly	AGT Ser	GGT Gly	GCA Ala	1728
				565					570					575		
ACA Thr	GGG Gly	TTC Phe	TCT Ser	GCA Ala	ACT Thr	GTT Val	GCA Ala	ACG Thr	GAG Glu	GTT Val	AAT Asn	ACC Thr	TCA Ser	ATT Ile	CAA Gln	1776
			580					585					590			
ATC Ile	CGT Arg	TCT Ser	GAC Asp	AGT Ser	CCT Pro	ACC Thr	GGA Gly	ACT Thr	CTA Leu	CTT Leu	GGT Gly	ACC Thr	TTA Leu	TAT Tyr	GTA Val	1824
		595					600					605				
AGT Ser	TCT Ser	ACC Thr	GGC Gly	AGC Ser	TGG Trp	AAT Asn	ACA Thr	TAT Tyr	CAA Gln	ACC Thr	GTA Val	TCT Ser	ACA Thr	AAC Asn	ATC Ile	1872
		610				615					620					
AGC Ser	AAA Lys	ATT Ile	ACC Thr	GGC Gly	GTT Val	CAT His	GAT Asp	ATT Ile	GTA Val	TTG Leu	GTA Val	TTC Phe	TCA Ser	GGT Gly	CCA Pro	1920
					630					635					640	
GTC Val	AAT Asn	GTG Val	GAC Asp	AAC Asn	TTC Phe	ATA Ile	TTT Phe	AGC Ser	AGA Arg	AGT Ser	TCA Ser	CCA Pro	GTG Val	CCT Pro	GCA Ala	1968
				645					650					655		
CCT Pro	GGT Gly	GAT Asp	AAC Asn	ACA Thr	AGA Arg	GAC Asp	GCA Ala	TAT Tyr	TCT Ser	ATC Ile	ATT Ile	CAG Gln	GCC Ala	GAG Glu	GAT Asp	2016
			660					665					670			
TAT Tyr	GAC Asp	AGC Ser	AGT Ser	TAT Tyr	GGT Gly	CCC Pro	AAC Asn	CTT Leu	CAA Gln	ATC Ile	TTT Phe	AGC Ser	TTA Leu	CCA Pro	GGT Gly	2064
		675					680					685				
GGT Gly	GGC Gly	AGC Ser	GCC Ala	ATT Ile	GGC Gly	TAT Tyr	ATT Ile	GAA Glu	AAT Asn	GGT Gly	TAT Tyr	TCC Ser	ACT Thr	ACC Thr	TAT Tyr	2112
		690				695					700					
AAA Lys	AAT Asn	ATT Ile	GAT Asp	TTT Phe	GGT Gly	GAC Asp	GGC Gly	GCA Ala	ACG Thr	TCC Ser	GTA Val	ACA Thr	GCA Ala	AGA Arg	GTA Val	2160
					710					715					720	

GCT ACC CAG AAT GCT ACT ACC ATT CAG GTA AGA TTG GGA AGT CCA TCG	2208
Ala Thr Gln Asn Ala Thr Thr Ile Gln Val Arg Leu Gly Ser Pro Ser	
725 730 735	
GGT ACA TTA CTT GGA ACA ATT TAC GTG GGG TCC ACA GGA AGC TTT GAT	2256
Gly Thr Leu Leu Gly Thr Ile Tyr Val Gly Ser Thr Gly Ser Phe Asp	
740 745 750	
ACT TAT AGG GAT GTA TCC GCT ACC ATT AGT AAT ACT GCG GGT GTA AAA	2304
Thr Tyr Arg Asp Val Ser Ala Thr Ile Ser Asn Thr Ala Gly Val Lys	
755 760 765	
GAT ATT GTT CTT GTA TTC TCA GGT CCT GTT AAT GTT GAC TGG	2346
Asp Ile Val Leu Val Phe Ser Gly Pro Val Asn Val Asp Trp	
770 775 780	

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 782 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Met	Lys	Gln	Gln	Lys	Arg	Leu	Tyr	Ala	Arg	Leu	Leu	Thr	Leu	Leu	Phe	1	5	10	15
Ala	Leu	Ile	Phe	Leu	Leu	Pro	His	Ser	Ala	Ala	Ala	Ala	Ala	Asn	Leu	20	25	30	
Asn	Gly	Thr	Leu	Met	Gln	Tyr	Phe	Glu	Trp	Tyr	Met	Pro	Asn	Asp	Gly	35	40	45	
Gln	His	Trp	Lys	Arg	Leu	Gln	Asn	Asp	Ser	Ala	Tyr	Leu	Ala	Glu	His	50	55	60	
Gly	Ile	Thr	Ala	Val	Trp	Ile	Pro	Pro	Ala	Tyr	Lys	Gly	Thr	Ser	Gln	65	70	75	80
Ala	Asp	Val	Gly	Tyr	Gly	Ala	Tyr	Asp	Leu	Tyr	Asp	Leu	Gly	Glu	Phe	85	90	95	
His	Gln	Lys	Gly	Thr	Val	Arg	Thr	Lys	Tyr	Gly	Thr	Lys	Gly	Glu	Leu	100	105	110	
Gln	Ser	Ala	Ile	Lys	Ser	Leu	His	Ser	Arg	Asp	Ile	Asn	Val	Tyr	Gly	115	120	125	
Asp	Val	Val	Ile	Asn	His	Lys	Gly	Gly	Ala	Asp	Ala	Thr	Glu	Asp	Val	130	135	140	
Thr	Ala	Val	Glu	Val	Asp	Pro	Ala	Asp	Arg	Asn	Arg	Val	Ile	Ser	Gly	145	150	155	160
Glu	His	Leu	Ile	Lys	Ala	Trp	Thr	His	Phe	His	Phe	Pro	Gly	Ala	Gly	165	170	175	
Ser	Thr	Tyr	Ser	Asp	Phe	Lys	Trp	His	Trp	Tyr	His	Phe	Asp	Gly	Thr	180	185	190	
Asp	Trp	Asp	Glu	Ser	Arg	Lys	Leu	Asn	Arg	Ile	Tyr	Lys	Phe	Gln	Gly	195	200	205	
Lys	Ala	Trp	Asp	Trp	Glu	Val	Ser	Asn	Glu	Asn	Gly	Asn	Tyr	Asp	Tyr	210	215	220	

Leu Met Tyr Ala Asp Ile Asp Tyr Asp His Pro Asp Val Ala Ala Glu
 225 230 235 240
 Ile Lys Arg Trp Gly Thr Trp Tyr Ala Asn Glu Leu Gln Leu Asp Gly
 245 250 255
 Asn Arg Leu Asp Ala Val Lys His Ile Lys Phe Ser Phe Leu Arg Asp
 260 265 270
 Trp Val Asn His Val Arg Glu Lys Thr Gly Lys Glu Met Phe Thr Val
 275 280 285
 Ala Glu Tyr Trp Gln Asn Asp Leu Gly Ala Leu Glu Asn Tyr Leu Asn
 290 295 300
 Lys Thr Asn Phe Asn His Ser Val Phe Asp Val Pro Leu His Tyr Gln
 305 310 315 320
 Phe His Ala Ala Ser Thr Gln Gly Gly Gly Tyr Asp Met Arg Lys Leu
 325 330 335
 Leu Asn Gly Thr Val Val Ser Lys His Pro Leu Lys Ser Val Thr Phe
 340 345 350
 Val Asp Asn His Asp Thr Gln Pro Gly Gln Ser Leu Glu Ser Thr Val
 355 360 365
 Gln Thr Trp Phe Lys Pro Leu Ala Tyr Ala Phe Ile Leu Thr Arg Glu
 370 375 380
 Ser Gly Tyr Pro Gln Val Phe Tyr Gly Asp Met Tyr Gly Thr Lys Gly
 385 390 395 400
 Asp Ser Gln Arg Glu Ile Pro Ala Leu Lys His Lys Ile Glu Pro Ile
 405 410 415
 Leu Lys Ala Arg Lys Gln Tyr Ala Tyr Gly Ala Gln His Asp Tyr Phe
 420 425 430
 Asp His His Asp Ile Val Gly Trp Thr Arg Glu Gly Asp Ser Ser Val
 435 440 445
 Ala Asn Ser Gly Leu Ala Ala Leu Ile Thr Asp Gly Pro Gly Gly Ala
 450 455 460
 Lys Arg Met Tyr Val Gly Arg Gln Asn Ala Gly Glu Thr Trp His Asp
 465 470 475 480
 Ile Thr Gly Asn Arg Ser Glu Pro Val Val Ile Asn Ser Glu Gly Trp
 485 490 495
 Gly Glu Phe His Val Asn Gly Gly Ser Val Ser Ile Tyr Val Gln Arg
 500 505 510
 Ser Pro Thr Pro Ala Pro Ser Gln Ser Pro Ile Arg Arg Asp Ala Phe
 515 520 525
 Ser Ile Ile Glu Ala Glu Glu Tyr Asn Ser Thr Asn Ser Ser Thr Leu
 530 535 540
 Gln Val Ile Gly Thr Pro Asn Asn Gly Arg Gly Ile Gly Tyr Ile Glu
 545 550 555 560
 Asn Gly Asn Thr Val Thr Tyr Ser Asn Ile Asp Phe Gly Ser Gly Ala
 565 570 575
 Thr Gly Phe Ser Ala Thr Val Ala Thr Glu Val Asn Thr Ser Ile Gln
 580 585 590

65

Ile Arg Ser Asp Ser Pro Thr Gly Thr Leu Leu Gly Thr Leu Tyr Val
 595 600 605
 Ser Ser Thr Gly Ser Trp Asn Thr Tyr Gln Thr Val Ser Thr Asn Ile
 610 615 620
 Ser Lys Ile Thr Gly Val His Asp Ile Val Leu Val Phe Ser Gly Pro
 625 630 635 640
 Val Asn Val Asp Asn Phe Ile Phe Ser Arg Ser Ser Pro Val Pro Ala
 645 650 655
 Pro Gly Asp Asn Thr Arg Asp Ala Tyr Ser Ile Ile Gln Ala Glu Asp
 660 665 670
 Tyr Asp Ser Ser Tyr Gly Pro Asn Leu Gln Ile Phe Ser Leu Pro Gly
 675 680 685
 Gly Gly Ser Ala Ile Gly Tyr Ile Glu Asn Gly Tyr Ser Thr Thr Tyr
 690 695 700
 Lys Asn Ile Asp Phe Gly Asp Gly Ala Thr Ser Val Thr Ala Arg Val
 705 710 715 720
 Ala Thr Gln Asn Ala Thr Thr Ile Gln Val Arg Leu Gly Ser Pro Ser
 725 730 735
 Gly Thr Leu Leu Gly Thr Ile Tyr Val Gly Ser Thr Gly Ser Phe Asp
 740 745 750
 Thr Tyr Arg Asp Val Ser Ala Thr Ile Ser Asn Thr Ala Gly Val Lys
 755 760 765
 Asp Ile Val Leu Val Phe Ser Gly Pro Val Asn Val Asp Trp
 770 775 780

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6136 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

TTTGACAGCT TATCATCGAC TGCACGGTGC ACCAATGCTT CTGGCGTCAG GCAGCCATCG	60
GAAGCTGTGG TATGGCTGTG CAGGTCGTAA ATCACTGCAT AATTCGTGTC GCTCAAGGCG	120
CACTCCCGTT CTGGATAATG TTTTGTGCGC CGACATCATA ACGGTTCTGG CAAATATTCT	180
GAAATGAGCT GTTGACAATT AATCATCGGC TCGTATAATG TGTGGAATTG TGAGCGGATA	240
ACAATTTTAC ACAGGAAACA GAATTGATCC ATAATACTAAT AATCTAGTAA TAATTTTGTT	300
TAACTTTAAAG AAGGAGATAT ATCCATGGAT CCTAGGACCA CGCCCGCACC CGGCCACCCG	360
GCCCGCGGCG CCCGCACCGC TCTGCGCAGC ACGCTCGCCG CCGCGGCGGC GACGCTCGTC	420
GTGCGCGCCA CGGTCTGTCT GCCCGCCAG GCGCTAGCG AATTCGCAA TCTTAATGGG	480
ACGCTGATGC AGTATTTTGA ATGGTACATG CCCAATGACG GCCAACATTG GAGGCGTTTG	540
CAAACGACT CGGCATATTT GGCTGAACAC GGTATTACTG CCGTCTGGAT TCCCCCGGCA	600
TATAAGGGAA CGAGCCAAGC GGATGTGGGC TACGGTGCTT ACGACCTTAA TGATTTAGGG	660

GAGTTTCATC AAAAAGGGAC GGTTCGGACA AAGTACGGCA CAAAAGGAGA GCTGCAATCT	720
GCGATCAAAA GTCTTCATTC CCGCGACATT AACGTTTACG GGGATGTGGT CATCAACCAC	780
AAAGGCGGCG CTGATGCGAC CGAAGATGTA ACCGCGGTTG AAGTCGATCC CGCTGACCGC	840
AACCGCGTAA TCTCAGGAGA ACACCTAATT AAAGCCTGGA CACATTTTCA TTTTCCGGGG	900
CGCGGCAGCA CATAAGCGA TTTTAAATGG CATTTGGTACC ATTTTGACGG AACCGATTGG	960
GACGAGTCCC GAAAGCTGAA CCGCATCTAT AAGTTTCAAG GAAAGGCTTG GGATTGGGAA	1020
GTTTCCAATG AAAACGGCAA CTATGATTAT TTGATGTATG CCGACATCGA TTATGACCAT	1080
CCTGATGTCG CAGCAGAAAT TAAGAGATGG GGCACCTGGT ATGCCAATGA ACTGCAATTG	1140
GACGGTTTCC GTCTTGATGC TGTCAAACAC ATTAAATTTT CTTTTTTGCG GGATTGGGTT	1200
AATCATGTCA GGGAAAAAAC GGGGAAGGAA ATGTTTACGG TAGCTGAATA TTGGCAGAAT	1260
GACTTGGGCG CGCTGGAAAA CTATTTGAAC AAAACAAATT TTAATCATTC AGTGTTTGAC	1320
GTGCCGCTTC ATTATCAGTT CCATGCTGCA TCGACACAGG GAGGCGGCTA TGATATGAGG	1380
AAATTGCTGA ACGGTACGGT CGTTTCCAAG CATCCGTGA AATCGGTTAC ATTTGTGCGAT	1440
AACCATGATA CACAGCCGGG GCAATCGCTT GAGTCGACTG TCCAAACATG GTTTAAGCCG	1500
CTTGCTTACG CTTTTATTCT CACAAGGGAA TCTGGATACC CTCAGGTTTT CTACGGGGAT	1560
ATGTACGGGA CGAAAGGAGA CTCCCAGCGC GAAATTCCTG CTTGAAACA CAAAATTGAA	1620
CCGATCTTAA AAGCGAGAAA ACAGTATGCG TACGGAGCAC AGCATGATTA TTTCGACCAC	1680
CATGACATTG TCGGCTGGAC AAGGGAAGGC GACAGCTCGG TTGCAAATTC AGGTTTGGCG	1740
GCATTAATAA CAGACGGACC CGGTGGGGCA AAGCGAATGT ATGTCGGCCG GCAAAACGCC	1800
GGTGAGACAT GGCATGACAT TACCGGAAAC CGTTCGGAGC CGGTTGTCAT CAATTCGGAA	1860
GGCTGGGGAG AGTTTCACGT AAACGGCGGG TCGGTTTCAA TTTATGTTCA AAGAAGGCCT	1920
CCAACCCCA CTAGTCCGAG CGTCCCAGC GGTGCACTG CTGAGAGGTG GGCTCAGTGC	1980
GGCGGCAATG GCTGGAGCGG CTGCACCACC TGCGTCGCTG GCAGCACTTG CACGAAGATT	2040
AATGACTGGT ACCATCAGTG CCTGTAAGCT TATTATATTA CTAATTAATT GGGGACCCTA	2100
GAGGTCCCCT TTTTATTTT AGCTTCACGC TGCCGCAAGC ACTCAGGGCG CAAGGGCTGC	2160
TAAAGGAAGC GGAACACGTA GAAAGCCAGT CCGCAGAAAC GGTGCTGACC CCGGATGAAT	2220
GTCAGCTACT GGGCTATCTG GACAAGGGAA AACGCAAGCG CAAAGAGAAA GCAGGTAGCT	2280
TGCAGTGGGC TTACATGGCG ATAGCTAGAC TGGGCGGTTT TATGGACAGC AAGCGAACCG	2340
GAATTGCCAG CTGGGGCGCC CTCTGGTAAG GTTGGGAAGC CCTGCAAAGT AACTGGATG	2400
GCTTTCTTGC CGCCAAGGAT CTGATGGCGC AGGGGATCAA GATCTGATCA AGAGACAGGA	2460
TGAGGATCGT TTCGCATGAT TGAACAAGAT GGATTGCACG CAGGTTCTCC GGCCGCTTGG	2520
GTGGAGAGGC TATTCGGCTA TGAAGGGCA CAACAGACAA TCGGCTGCTC TGATGCCGCC	2580
GTGTTCCGGC TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG TCAAGACCGA CCTGTCCGGT	2640
GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT GGCTGGCCAC GACGGGCGTT	2700

CCTTGCGCAG CTGTGCTCGA CGTTGTCCT GAAGCGGGAA GGGACTGGCT GCTATTGGGC 2760
GAAGTGCCGG GGCAGGATCT CCTGTCATCT CACCTTGCTC CTGCCGAGAA AGTATCCATC 2820
ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG CTACCTGCCC ATTCGACCAC 2880
CAAGCGAAAC ATCGCATCGA GCGAGCACGT ACTCGGATGG AAGCCGGTCT TGTCGATCAG 2940
GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG AACTGTTCGC CAGGCTCAAG 3000
GCGCGCATGC CCGACGGCGA GGATCTCGTC GTGACACATG GCGATGCCTG CTTGCCGAAT 3060
ATCATGGTGG AAAATGGCCG CTTTCTGGA TTCATCGACT GTGGCCGGCT GGGTGTGGCG 3120
GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG CTGAAGAGCT TGGCGGCGAA 3180
TGGGCTGACC GCTTCCTCGT GCTTTACGGT ATCGCCGCTC CCGATTGCGA GCGCATCGCC 3240
TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT GGGGTTCGAA ATGACCGACC 3300
AAGCGACGCC CAACCTGCCA TCACGAGATT TCGATTCCAC CGCCGCCTTC TATGAAAGGT 3360
TGGGCTTCGG AATCGTTTTT CCGGACGCCG GCTGGATGAT CCTCCAGCGC GGGGATCTCA 3420
TGCTGGAGTT CTTGCCCCAC CCCAAAAGGA TCTAGGTGAA GATCCTTTTT GATAATCTCA 3480
TGACCAAAAT CCCTTAACGT GAGTTTTCTG TCCACTGAGC GTCAGACCCC GTAGAAAAGA 3540
TCAAAGGATC TTCTTGAGAT CCTTTTTTTC TGCGCGTAAT CTGCTGCTTG CAAACAAAAA 3600
AACCACCGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT CTTTTTCCGA 3660
AGGTAAC TGG CTT CAGCAGA GCGCAGATAC CAAATACTGT CCTTCTAGTG TAGCCGTAGT 3720
TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCCTACATA CCTCGCTCTG CTAATCCTGT 3780
TACCAGTGGC TGCTGCCAGT GGCGATAAGT CGTGTCTTAC CGGGTTGGAC TCAAGACGAT 3840
AGTTACCGGA TAAGGCGCAG CGGTCGGGCT GAACGGGGGG TTCGTGCACA CAGCCCAGCT 3900
TGGAGCGAAC GACCTACACC GAACTGAGAT ACCTACAGCG TGAGCTATGA GAAAGCGCCA 3960
CGCTTCCCGA AGGGAGAAAG GCGGACAGGT ATCCGGTAAG CGGCAGGGTC GGAACAGGAG 4020
AGCGCACGAG GGAGCTTCCA GGGGGAACG CCTGGTATCT TTATAGTCCT GTCGGGTTTC 4080
GCCACCTCTG ACTTGAGCGT CGATTTTTGT GATGCTCGTC AGGGGGGCGG AGCCTATGGA 4140
AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGGCCTT TTGCTGGCCT TTTGCTCACA 4200
TGTTCTTTCC TGCCTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG 4260
CTGATACCGC TCGCCGACG CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG 4320
AAGAGCGCCT GATGCGGTAT TTTCTCCTTA CGCATCTGTG CGGTATTTCA CACCGCATAT 4380
GCAGATATTT TGTTAAAATT CGCGTTAAAT TTTTGTTAA TCAGCTCATT TTTTAACCAA 4440
TAGGCCGAAA TCGGCAAAAT CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGGTTGAGT 4500
GTTGTTCCAG TTTGGAACAA GAGTCCACTA TTAAAGAACG TGGACTCCAA CGTCAAAGGG 4560
CGAAAAACCG TCTATCAGGG CGATGGCCCA CTACGTGAAC CATCACCTA ATCAAGTTTT 4620
TTGGGGTCGA GGTGCCGTAA AGCACTAAAT CGGAACCCTA AAGGGAGCCC CCGATTTAGA 4680
GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG 4740

```

GGCGCTAGGG CGCTGGCAAG TGTAGCGGTC ACGCTGCGCG TAACCACCAC ACCCGCCGCG 4800
CTTAATGCGC CGCTACAGGG CGCGTCAGGT GGCACTTTTC GGGGAAATGT GCGCGGAACC 4860
CCTATTTGTT TATTTTCTA AATACATTCA AATATGTATC CGCTCATGAG ACAATAACCC 4920
TGCTGCATTT ACGTTGACAC CATCGAATGG TGCAAAACCT TTCGCGGTAT GGCATGATAG 4980
CGCCCCGAAG AGAGTCAATT CAGGGTGGTG AATGTGAAAC CAGTAACGTT ATACGATGTC 5040
GCAGAGTATG CCGGTGTCTC TTATCAGACC GTTTCCTCGG TGGTGAACCA GGCCAGCCAC 5100
GTTTCTGCGA AAACGCGGGA AAAAGTGGAA GCGGCGATGG CGGAGCTGAA TTACATTCCC 5160
AACCGCGTGG CACAACAACCT GGCGGGCAAA CAGTCGTTGC TGATTGGCGT TGCCACCTCC 5220
AGTCTGGCCC TGCACGCGCC GTCGCAAATT GTCGCGGCGA TTAAATCTCG CGCCGATCAA 5280
CTGGGTGCCA GCGTGGTGGT GTCGATGGTA GAACGAAGCG GCGTCGAAGC CTGTAAAGCG 5340
GCGGTGCACA ATCTTCTCGC GCAACGCGTC AGTGGGCTGA TCATTAATA TCCGCTGGAT 5400
GACCAGGATG CCATTGCTGT GGAAGCTGCC TGACTAATG TTCCGGCGTT ATTTCTTGAT 5460
GTCTCTGACC AGACACCCAT CAACAGTATT ATTTTCTCCC ATGAAGACGG TACGCGACTG 5520
GGCGTGGAGC ATCTGGTCGC ATTGGGTCAC CAGCAAATCG CGCTGTTAGC GGGCCCATT 5580
AGTTCTGTCT CGGCGCGTCT GCGTCTGGCT GGCTGGCATA AATATCTCAC TCGCAATCAA 5640
ATTCAGCCGA TAGCGGAACG GGAAGGCGAC TGGAGTGCCA TGTCGGGTTT TCAACAAACC 5700
ATGCAAATGC TGAATGAGGG CATCGTTCCC ACTGCGATGC TGGTTGCCAA CGATCAGATG 5760
GCGCTGGGCG CAATGCGCGC CATTACCGAG TCCGGGCTGC GCGTTGGTGC GGATATCTCG 5820
GTAGTGGGAT ACGACGATAC CGAAGACAGC TCATGTTATA TCCCGCCGTT AACCACCATC 5880
AAACAGGATT TTCGCTGCT GGGGCAAACC AGCGTGACC GCTTGCTGCA ACTCTCTCAG 5940
GGCCAGGCGG TGAAGGGCAA TCAGCTGTTG CCCGTCTCAC TGGTGAAAAG AAAAACCACC 6000
CTGGCGCCCA ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCAAT AATGCAGCTG 6060
GCACGACAGG TTTCCCGACT GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA 6120
GCGCGAATTG ATCTGG 6136

```

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(ix) FEATURE:

(A) NAME/KEY: misc-feature:

(B) OTHER INFORMATION: /desc = "Primer 14"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AGGTCTACTA GTCCCGGCTG CCGCGTCGAC

30

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 base pairs

69

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: other nucleic acid
- (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 15"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CCGATTAAAG CTTATTAGCT AGCACGGAAT TCCGTGGGGC TGGTCGTCGG CAC

53

- (2) INFORMATION FOR SEQ ID NO:29:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 16"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TCATGAGCCA TGGCTAGCGC AAATCTTAAT GGGACGCTGA TG

42

- (2) INFORMATION FOR SEQ ID NO:30:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: other nucleic acid
 - (ix) FEATURE:
 - (A) NAME/KEY: misc-feature:
 - (B) OTHER INFORMATION: /desc = "Primer 17"
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATGACTAAGC TTAATTACTT AGTGATGGTG ATGGTGATGA CTAGTTCTTT GAACATAAAT TGAAACCGA

69

- (2) INFORMATION FOR SEQ ID NO: 31:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1959 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: cDNA
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..1959
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

ATG GAT CCT AGG ACC ACG CCC GCA CCC GGC CAC CCG GCC CGC GGC GCC 48
 Met Asp Pro Arg Thr Thr Pro Ala Pro Gly His Pro Ala Arg Gly Ala
 1 5 10 15

CGC ACC GCT CTG CGC ACG ACG CTC GCC GCC GCG GCG GCG ACG CTC GTC 96
 Arg Thr Ala Leu Arg Thr Thr Leu Ala Ala Ala Ala Thr Leu Val
 20 25 30

GTC GGC GCC ACG GTC GTG CTG CCC GCC CAG GCC GCT AGT CCC GGC TGC 144
 Val Gly Ala Thr Val Val Leu Pro Ala Gln Ala Ala Ser Pro Gly Cys
 35 40 45

CGC GTC GAC TAC GCC GTC ACC AAC CAG TGG CCC GGC GGC TTC GGC GCC 192
 Arg Val Asp Tyr Ala Val Thr Asn Gln Trp Pro Gly Gly Phe Gly Ala

70

50	55	60	
AAC GTC ACG ATC ACC AAC CTC GGC GAC CCC GTC TCG TCG TGG AAG CTC Asn Val Thr Ile Thr Asn Leu Gly Asp Pro Val Ser Ser Trp Lys Leu 65 70 75 80			240
GAC TGG ACC TAC ACC GCA GGC CAG CGG ATC CAG CAG CTG TGG AAC GGC Asp Trp Thr Tyr Thr Ala Gly Gln Arg Ile Gln Gln Leu Trp Asn Gly 85 90 95			288
ACC GCG TCG ACC AAC GGC GGC CAG GTC TCC GTC ACC AGC CTG CCC TGG Thr Ala Ser Thr Asn Gly Gly Gln Val Ser Val Thr Ser Leu Pro Trp 100 105 110			336
AAC GGC AGC ATC CCG ACC GGC GGC ACG GCG TCG TTC GGG TTC AAC GGC Asn Gly Ser Ile Pro Thr Gly Gly Thr Ala Ser Phe Gly Phe Asn Gly 115 120 125			384
TCG TGG GCC GGG TCC AAC CCG ACG CCG GCG TCG TTC TCG CTC AAC GGC Ser Trp Ala Gly Ser Asn Pro Thr Pro Ala Ser Phe Ser Leu Asn Gly 130 135 140			432
ACC ACC TGC ACG GGC ACC GTG CCG ACG ACC AGC CCC ACG GAA TTC CGT Thr Thr Cys Thr Gly Thr Val Pro Thr Thr Ser Pro Thr Glu Phe Arg 145 150 155 160			480
GCT AGC GCA AAT CTT AAT GGG ACG CTG ATG CAG TAT TTT GAA TGG TAC Ala Ser Ala Asn Leu Asn Gly Thr Leu Met Gln Tyr Phe Glu Trp Tyr 165 170 175			528
ATG CCC AAT GAC GGC CAA CAT TGG AAG CGC TTG CAA AAC GAC TCG GCA Met Pro Asn Asp Gly Gln His Trp Lys Arg Leu Gln Asn Asp Ser Ala 180 185 190			576
TAT TTG GCT GAA CAC GGT ATT ACT GCC GTC TGG ATT CCC CCG GCA TAT Tyr Leu Ala Glu His Gly Ile Thr Ala Val Trp Ile Pro Pro Ala Tyr 195 200 205			624
AAG GGA ACG AGC CAA GCG GAT GTG GGC TAC GGT GCT TAC GAC CTT TAT Lys Gly Thr Ser Gln Ala Asp Val Gly Tyr Gly Ala Tyr Asp Leu Tyr 210 215 220			672
GAT TTA GGG GAG TTT CAT CAA AAA GGG ACG GTT CGG ACA AAG TAC GGC Asp Leu Gly Glu Phe His Gln Lys Gly Thr Val Arg Thr Lys Tyr Gly 225 230 235 240			720
ACA AAA GGA GAG CTG CAA TCT GCG ATC AAA AGT CTT CAT TCC CGC GAC Thr Lys Gly Glu Leu Gln Ser Ala Ile Lys Ser Leu His Ser Arg Asp 245 250 255			768
ATT AAC GTT TAC GGG GAT GTG GTC ATC AAC CAC AAA GGC GGC GCT GAT Ile Asn Val Tyr Gly Asp Val Val Ile Asn His Lys Gly Gly Ala Asp 260 265 270			816
GCG ACC GAA GAT GTA ACC GCG GTT GAA GTC GAT CCC GCT GAC CGC AAC Ala Thr Glu Asp Val Thr Ala Val Glu Val Asp Pro Ala Asp Arg Asn 275 280 285			864
CGC GTA ATT TCA GGA GAA CAC TTA ATT AAA GCC TGG ACA CAT TTT CAT Arg Val Ile Ser Gly Glu His Leu Ile Lys Ala Trp Thr His Phe His 290 295 300			912
TTT CCG GGG CGC GGC AGC ACA TAC AGC GAT TTT AAA TGG CAT TGG TAC Phe Pro Gly Arg Gly Ser Thr Tyr Ser Asp Phe Lys Trp His Trp Tyr 305 310 315 320			960
CAT TTT GAC GGA ACC GAT TGG GAC GAG TCC CGA AAG CTG AAC CGC ATC His Phe Asp Gly Thr Asp Trp Asp Glu Ser Arg Lys Leu Asn Arg Ile			1008

																1056
TAT Tyr	AAG Lys	TTT Phe	CAA Gln 340	GGA Gly	AAG Lys	GCT Ala	TGG Trp	GAT Asp 345	TGG Trp	GAA Glu	GTT Val	TCC Ser	AAT Asn 350	GAA Glu	AAC Asn	1056
GGC Gly	AAC Asn	TAT Tyr 355	GAT Asp	TAT Tyr	TTG Leu	ATG Met	TAT Tyr 360	GCC Ala	GAC Asp	ATC Ile	GAT Asp 365	TAT Tyr	GAT Asp	CAT His	CCT Pro	1104
GAT Asp	GTC Val 370	GCA Ala	GCA Ala	GAA Glu	ATT Ile	AAG Lys 375	AGA Arg	TGG Trp	GGC Gly	ACT Thr 380	TGG Trp	TAT Tyr	GCC Ala	AAT Asn	GAA Glu	1152
CTG Leu 385	CAA Gln	TTG Leu	GAC Asp	GGT Gly	TTC Phe 390	CGT Arg	CTT Leu	GAT Asp	GCT Ala	GTC Val 395	AAA Lys	CAC His	ATT Ile	AAA Lys	TTT Phe 400	1200
TCT Ser	TTT Phe	TTG Leu	CGG Arg	GAT Asp 405	TGG Trp	GTT Val	AAT Asn	CAT His	GTC Val 410	AGG Arg	GAA Glu	AAA Lys	ACG Thr	GGG Gly 415	AAG Lys	1248
GAA Glu	ATG Met	TTT Phe	ACG Thr 420	GTA Val	GCT Ala	GAA Glu	TAT Tyr	TGG Trp 425	CAG Gln	AAT Asn	GAC Asp	TTG Leu	GGC Gly 430	GCG Ala	CTG Leu	1296
GAA Glu	AAC Asn	TAT Tyr 435	TTG Leu	AAC Asn	AAA Lys	ACA Thr	AAT Asn 440	TTT Phe	AAT Asn	CAT His	TCA Ser	GTG Val 445	TTT Phe	GAC Asp	GTG Val	1344
CCG Pro	CTT Leu 450	CAT His	TAT Tyr	CAG Gln	TTC Phe	CAT His 455	GCT Ala	GCA Ala	TCG Ser	ACA Thr 460	CAG Gln	GGA Gly	GGC Gly	GGC Gly	TAT Tyr	1392
GAT Asp 465	ATG Met	AGG Arg	AAA Lys	TTG Leu	CTG Leu 470	AAC Asn	GGT Gly	ACG Thr	GTC Val 475	GTT Val 475	TCC Ser	AAG Lys	CAT His	CCG Pro	TTG Leu 480	1440
AAA Lys	GCG Ala	GTT Val	ACA Thr 485	TTT Phe	GTC Val	GAT Asp	AAC Asn	CAT His 490	GAT Asp 490	ACA Thr	CAG Gln	CCG Pro	GGG Gly	CAA Gln 495	TCG Ser	1488
CTT Leu	GAG Glu	TCG Ser 500	ACT Thr	GTC Val	CAA Gln	ACA Thr	TGG Trp	TTT Phe 505	AAG Lys	CCG Pro	CTT Leu	GCT Ala 510	TAC Tyr	GCT Ala	TTT Phe	1536
ATT Ile	CTC Leu	ACA Thr 515	AGG Arg	GAA Glu	TCT Ser	GGA Gly	TAC Tyr 520	CCT Pro	CAG Gln	GTT Val	TTC Phe 525	TAC Tyr	GGG Gly	GAT Asp	ATG Met	1584
TAC Tyr	GGG Gly 530	ACG Thr	AAA Lys	GGA Gly	GAC Asp	TCC Ser 535	CAG Gln	CGC Arg	GAA Glu	ATT Ile 540	CCT Pro 540	GCC Ala	TTG Leu	AAA Lys	CAC His	1632
AAA Lys 545	ATT Ile	GAA Glu	CCG Pro	ATC Ile	TTA Leu 550	AAA Lys	GCG Ala	AGA Arg	AAA Lys 555	CAG Gln	TAT Tyr	GCG Ala	TAC Tyr	GGA Gly	GCA Ala 560	1680
CAG Gln	CAT His	GAT Asp	TAT Tyr 565	TTC Phe	GAC Asp	CAC His	CAT His	GAC Asp 570	ATT Ile	GTC Val	GGC Gly	TGG Trp	ACA Thr	AGG Arg 575	GAA Glu	1728
GGC Gly	GAC Asp	AGC Ser 580	TCG Ser	GTT Val	GCA Ala	AAT Asn	TCA Ser	GGT Gly 585	TTG Leu	GCG Ala	GCA Ala	TTA Leu 590	ATA Ile	ACA Thr	GAC Asp	1776
GGA Gly	CCC Pro	GGT Gly	GGG Gly	GCA Ala	AAG Lys	CGA Arg	ATG Met	TAT Tyr	GTC Val	GGC Gly	CGG Arg	CAA Gln	AAC Asn	GCC Ala	GGT Gly	1824

595					600					605						
GAG	ACA	TGG	CAT	GAC	ATT	ACC	GGA	AAC	CGT	TCG	GAG	CCG	GTT	GTC	ATC	1872
Glu	Thr	Trp	His	Asp	Ile	Thr	Gly	Asn	Arg	Ser	Glu	Pro	Val	Val	Ile	
610					615					620						
AAT	TCG	GAA	GGC	TGG	GGA	GAG	TTT	CAC	GTA	AAC	GGC	GGG	TCG	GTT	TCA	1920
Asn	Ser	Glu	Gly	Trp	Gly	Glu	Phe	His	Val	Asn	Gly	Gly	Ser	Val	Ser	
625					630					635					640	
ATT	TAT	GTT	CAA	AGA	ACT	AGT	CAT	CAC	CAT	CAC	CAT	CAC				
Ile	Tyr	Val	Gln	Arg	Thr	Ser	His	His	His	His	His	His				
				645					650							

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 653 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Met	Asp	Pro	Arg	Thr	Thr	Pro	Ala	Pro	Gly	His	Pro	Ala	Arg	Gly	Ala	
1				5					10					15		
Arg	Thr	Ala	Leu	Arg	Thr	Thr	Leu	Ala	Ala	Ala	Ala	Ala	Thr	Leu	Val	
			20					25					30			
Val	Gly	Ala	Thr	Val	Val	Leu	Pro	Ala	Gln	Ala	Ala	Ser	Pro	Gly	Cys	
		35				40						45				
Arg	Val	Asp	Tyr	Ala	Val	Thr	Asn	Gln	Trp	Pro	Gly	Gly	Phe	Gly	Ala	
	50					55					60					
Asn	Val	Thr	Ile	Thr	Asn	Leu	Gly	Asp	Pro	Val	Ser	Ser	Trp	Lys	Leu	
	65				70					75					80	
Asp	Trp	Thr	Tyr	Thr	Ala	Gly	Gln	Arg	Ile	Gln	Gln	Leu	Trp	Asn	Gly	
				85					90					95		
Thr	Ala	Ser	Thr	Asn	Gly	Gly	Gln	Val	Ser	Val	Thr	Ser	Leu	Pro	Trp	
			100				105						110			
Asn	Gly	Ser	Ile	Pro	Thr	Gly	Gly	Thr	Ala	Ser	Phe	Gly	Phe	Asn	Gly	
		115				120						125				
Ser	Trp	Ala	Gly	Ser	Asn	Pro	Thr	Pro	Ala	Ser	Phe	Ser	Leu	Asn	Gly	
	130					135					140					
Thr	Thr	Cys	Thr	Gly	Thr	Val	Pro	Thr	Thr	Ser	Pro	Thr	Glu	Phe	Arg	
	145				150					155					160	
Ala	Ser	Ala	Asn	Leu	Asn	Gly	Thr	Leu	Met	Gln	Tyr	Phe	Glu	Trp	Tyr	
			165						170					175		
Met	Pro	Asn	Asp	Gly	Gln	His	Trp	Lys	Arg	Leu	Gln	Asn	Asp	Ser	Ala	
			180					185					190			
Tyr	Leu	Ala	Glu	His	Gly	Ile	Thr	Ala	Val	Trp	Ile	Pro	Pro	Ala	Tyr	
		195				200						205				
Lys	Gly	Thr	Ser	Gln	Ala	Asp	Val	Gly	Tyr	Gly	Ala	Tyr	Asp	Leu	Tyr	
	210					215					220					
Asp	Leu	Gly	Glu	Phe	His	Gln	Lys	Gly	Thr	Val	Arg	Thr	Lys	Tyr	Gly	
	225				230					235					240	

Thr Lys Gly Glu Leu Gln Ser Ala Ile Lys Ser Leu His Ser Arg Asp
 245 250 255
 Ile Asn Val Tyr Gly Asp Val Val Ile Asn His Lys Gly Gly Ala Asp
 260 265 270
 Ala Thr Glu Asp Val Thr Ala Val Glu Val Asp Pro Ala Asp Arg Asn
 275 280 285
 Arg Val Ile Ser Gly Glu His Leu Ile Lys Ala Trp Thr His Phe His
 290 295 300
 Phe Pro Gly Arg Gly Ser Thr Tyr Ser Asp Phe Lys Trp His Trp Tyr
 305 310 315 320
 His Phe Asp Gly Thr Asp Trp Asp Glu Ser Arg Lys Leu Asn Arg Ile
 325 330 335
 Tyr Lys Phe Gln Gly Lys Ala Trp Asp Trp Glu Val Ser Asn Glu Asn
 340 345 350
 Gly Asn Tyr Asp Tyr Leu Met Tyr Ala Asp Ile Asp Tyr Asp His Pro
 355 360 365
 Asp Val Ala Ala Glu Ile Lys Arg Trp Gly Thr Trp Tyr Ala Asn Glu
 370 375 380
 Leu Gln Leu Asp Gly Phe Arg Leu Asp Ala Val Lys His Ile Lys Phe
 385 390 395 400
 Ser Phe Leu Arg Asp Trp Val Asn His Val Arg Glu Lys Thr Gly Lys
 405 410 415
 Glu Met Phe Thr Val Ala Glu Tyr Trp Gln Asn Asp Leu Gly Ala Leu
 420 425 430
 Glu Asn Tyr Leu Asn Lys Thr Asn Phe Asn His Ser Val Phe Asp Val
 435 440 445
 Pro Leu His Tyr Gln Phe His Ala Ala Ser Thr Gln Gly Gly Gly Tyr
 450 455 460
 Asp Met Arg Lys Leu Leu Asn Gly Thr Val Val Ser Lys His Pro Leu
 465 470 475 480
 Lys Ala Val Thr Phe Val Asp Asn His Asp Thr Gln Pro Gly Gln Ser
 485 490 495
 Leu Glu Ser Thr Val Gln Thr Trp Phe Lys Pro Leu Ala Tyr Ala Phe
 500 505 510
 Ile Leu Thr Arg Glu Ser Gly Tyr Pro Gln Val Phe Tyr Gly Asp Met
 515 520 525
 Tyr Gly Thr Lys Gly Asp Ser Gln Arg Glu Ile Pro Ala Leu Lys His
 530 535 540
 Lys Ile Glu Pro Ile Leu Lys Ala Arg Lys Gln Tyr Ala Tyr Gly Ala
 545 550 555 560
 Gln His Asp Tyr Phe Asp His His Asp Ile Val Gly Trp Thr Arg Glu
 565 570 575
 Gly Asp Ser Ser Val Ala Asn Ser Gly Leu Ala Ala Leu Ile Thr Asp
 580 585 590
 Gly Pro Gly Gly Ala Lys Arg Met Tyr Val Gly Arg Gln Asn Ala Gly
 595 600 605

74

Glu Thr Trp His Asp Ile Thr Gly Asn Arg Ser Glu Pro Val Val Ile
 610 615 620
 Asn Ser Glu Gly Trp Gly Glu Phe His Val Asn Gly Gly Ser Val Ser
 625 630 635 640
 Ile Tyr Val Gln Arg Thr Ser His His His His His His
 645 650

- (2) INFORMATION FOR SEQ ID NO:33:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 29 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 18"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CATATGGCTA GCGAATTCGC AAATCTTAAT GGGACGCTG

29

- (2) INFORMATION FOR SEQ ID NO:34:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 28 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 19"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AAGCTTACTA GTAGGCCTTC TTTGAACATA AATTGAAA

28

- (2) INFORMATION FOR SEQ ID NO:35:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 70 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 20"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCATGGGCTA GCCCTGAATT CAGGCCTCCA ACCCCCACTA GTCCGAGCGC TCCCAGCGGC
 TGCACTGCTG 70

- (2) INFORMATION FOR SEQ ID NO:36:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 32 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (ii) MOLECULE TYPE: other nucleic acid
 (ix) FEATURE:
 (A) NAME/KEY: misc-feature:
 (B) OTHER INFORMATION: /desc = "Primer 21"
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

AGCCTAAGCT TACAGGCACT GATGGTACCA GT

32

2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 12 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (ix) FEATURE:
 - (a) NAME/KEY: misc-feature
 - (d) OTHER INFORMATION: /desc = "Linker"
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Arg Pro Pro Thr Pro Thr Ser Pro Ser Ala Pro Ser
1 5 10

CLAIMS

1. A method for liquefying starch, wherein a starch substrate is treated in aqueous medium with a modified enzyme (enzyme hybrid)
5 which comprises an amino acid sequence of an α -amylase linked to an amino acid sequence comprising a carbohydrate-binding domain (CBD).
2. The method for liquefying starch according to claim 1, further
10 comprising a debranching enzyme.
3. The method according to claim 2, wherein the debranching enzyme is a modified debranching enzyme (enzyme hybrid) linked to an amino acid sequence comprising a carbohydrate-binding domain.
15
4. A method for saccharifying starch which has been subjected to a liquefaction process, wherein the reaction mixture after liquefaction is treated with a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of a debranching enzyme
20 linked to an amino acid sequence comprising a carbohydrate-binding domain (CBD).
5. The method according to claims 2, 3 or 4 wherein said debranching enzyme is an isoamylase or a pullulanase.
25
6. A method for saccharifying starch which has been subjected to a liquefaction process, wherein the reaction mixture after liquefaction is treated with a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of a glucoamylase linked
30 to an amino acid sequence comprising a carbohydrate-binding domain (CBD).
7. A method according to any one of the preceding claims, wherein said CBD is a CBD deriving from a cellulase, a xylanase, a
35 mannanase, an arabinofuranosidase, an acetylerase, a chitinase, a glucoamylase or a CGTase.

8. The use of a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of an α -amylase linked to an amino acid sequence comprising a carbohydrate-binding domain (CBD) in a process for liquefying starch.

5

9. The use of a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of a debranching enzyme linked to an amino acid sequence comprising a carbohydrate-binding domain (CBD) in a process for saccharifying starch which has been subjected to a
10 liquefaction process.

10. The use of a modified enzyme (enzyme hybrid) which comprises an amino acid sequence of a glucoamylase linked to an amino acid sequence comprising a carbohydrate-binding domain (CBD) in a
15 process for saccharifying starch which has been subjected to a liquefaction process.

11. An isolated DNA sequence encoding a hybrid enzyme with amylolytic activity comprising:

- 20 (a) a DNA sequence encoding an amylolytic activity;
(b) a DNA sequences encoding a CBD; and
(c) a DNA sequence or fragments thereof encoding the linker sequence shown in SEQ ID no. 21.

25 12. The isolated DNA sequence according to claims 11, wherein the amylolytic activity is an α -amylase activity, in particular a *Bacillus* α -amylase, especially the activity of Termamyl[®] or a variant thereof.

30 13. The isolated DNA sequence according to claims 11 or 12, wherein the CBD is the CBD of *Bacillus agaradherens* NCIMB No. 40482 alkaline cellulase Cel5A.

14. The isolated DNA sequence according to claim 13, encodes
35 the Termamyl[®]-linker-Cel5A-CBD encoded by plasmid pMB492 shown in SEQ ID No. 19.

15. The isolated DNA sequence according to claims 11 or 12, wherein the CDB is the CBD-dimer of *Clostridium stercorarium* (NCIMB 11754) XynA.

5

16. A DNA construct comprising the DNA sequence of any of claims 11 to 15 operably linked to one or more control sequences capable of directing the expression of the DNA sequence in a suitable expression host.

10

17. The DNA construct of claim 16, comprising a nucleotide sequence encoding the promoter selected from the group consisting of the promoter of the *Bacillus stearothermophilus* maltogenic amylase gene, the promoter of the *Bacillus*
15 *licheniformis* alpha-amylase gene, the promoter of the *Bacillus amyloliquefaciens* BAN~~XXXX~~ amylase gene, the promoter of the *Bacillus subtilis* alkaline protease gene, or the promoter of the *Bacillus pumilus* cellulase or xylosidase gene.

20 18. A recombinant expression vector comprising the DNA construct of claims 16 or 17, a promoter, and transcriptional and translational stop signals.

19. A host cell comprising the DNA construct of claims 16 or
25 17.

20. The cell of claim 19, wherein the cell is a *Bacillus* cell from a strain selected from the group consisting of *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. brevis*, *B. stearothermophilus*, *B. alkalophilus*, *B. amyloliquefaciens*, *B.*
30 *coagulans*, *B. circulans*, *B. lautus*, *B. megatherium*, *B. pumilus*, *B. thuringiensis* or *B. agaradherens*.

21. A method of producing a CBD/ hybrid enzyme, comprised
35 of culturing the cell of claims 19 or 20 under conditions permitting the production of the enzyme, and recovering the enzyme from the culture.

22. An isolated and purified CBD/enzyme hybrid encoded by the DNA sequence of any of claims 11 to 15.

5 23. The CBD/enzyme hybrid according to claim 22 being the hybrid enzyme shown in SEQ ID No. 20.

INTERNATIONAL SEARCH REPORT

International application No. -

PCT/DK 97/00448

A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 9/26, C12N 15/56, C07K 19/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N, C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, WPI

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Elsevier Science Ltd, Volume 12, 1994, Edward A. Bayer et al, "The cellulosome - a treasuretrove for biotechnology" --	1-23
Y	US 5496934 A (ODED SHOSEYOV ET AL), 5 March 1996 (05.03.96) --	1-23
Y	WO 9429460 A1 (MIDWEST RESEARCH INSTITUTE), 22 December 1994 (22.12.94) --	1-23
Y	WO 9623874 A1 (NOVO NORDISK A/S), 8 August 1996 (08.08.96) --	1-23

☒ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"B" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

4 February 1998

Date of mailing of the international search report

06 -02- 1998

Name and mailing address of the ISA/
Swedish Patent Office
Box 5055, S-102 42 STOCKHOLM
Facsimile No. +46 8 666 02 86

Authorized officer

Carl-Olof Gustafsson
Telephone No. +46 8 782 25 00

INTERNATIONAL SEARCH REPORT

International application No. -

PCT/DK 97/00448

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>Journal of Bacteriology, Volume 177, No 18, Sept 1995, Nathalie Sauvonnet et al, "Extracellular Secretion of Pullulanase Is Unaffected by Minor Sequence Changes but Is Usually Prevented by Adding Reporter Proteins to Its N- or C-Terminal End" page 5241 - page 5243</p> <p style="text-align: center;">-- -----</p>	2,3,5

INTERNATIONAL SEARCH REPORT
Information on patent family members

07/01/98

International application No. -

PCT/DK 97/00448

Patent document cited in search report			Publication date	Patent family member(s)		Publication date
US	5496934	A	05/03/96	AU	6634794 A	08/11/94
				CA	2160670 A	27/10/94
				CN	1125452 A	26/06/96
				EP	0695311 A	07/02/96
				FI	954888 A	13/12/95
				IL	109323 D	00/00/00
				JP	8509127 T	01/10/96
				NO	954074 A	27/11/95
				NZ	265537 A	26/05/97
				US	5670623 A	23/09/97
				WO	9424158 A	27/10/94

WO	9429460	A1	22/12/94	NONE		

WO	9623874	A1	08/08/96	AU	4483496 A	21/08/96
				CA	2211316 A	08/08/96
				EP	0808363 A	26/11/97
